

GYAN BHARATAM

International Conference on

Reclaiming India's Knowledge
Legacy Through Manuscript Heritage

11th-13th September 2025

WORKING GROUP REPORTS



संस्कृति मंत्रालय
Ministry of Culture
Government of India



Ministry of Culture
Government of India

GYAN BHARATAM

International Conference
on

Reclaiming India's Knowledge
Legacy Through Manuscript Heritage

WORKING GROUP REPORTS

Editors:

Ms. Amita Prasad Sarbhai
Additional Secretary, Ministry of Culture, Govt. of India

Dr. Sachchidanand Joshi
Member Secretary
Indira Gandhi National Centre for the Arts

Prof. (Dr.) Ramesh C. Gaur
Dean Administration
Indira Gandhi National Centre for the Arts

CHIEF PATRON

Shri Gajendra Singh Shekawat
Hon'ble Union Minister of Culture & Tourism

PATRON

Shri Vivek Aggarwal
Secretary
Ministry of Culture, Govt. of India

ORGANIZING COMMITTEE

Convenor
Ms. Amita Prasad Sarbhai
Additional Secretary, Ministry
of Culture, Govt. of India

Co-Convenor
Dr. Sachchidanand Joshi
Member Secretary
Indira Gandhi National Centre
for the Arts

Member
Prof. Alok Tripathi
Addl. Director General
Archaeological Survey of India

Member
Shri Samar Nanda
Joint Secretary, Ministry of
Culture, Govt. of India

Member
Prof. Ramesh C. Gaur
Dean Administration
Indira Gandhi National Centre
for the Arts

Member
Prof. Anirban Dash
Project Director
Gyan Bharatam

ADVISORY COMMITTEE

Prof. Chamu Krishna Shastri
Chairman Bharatiya Bhasha
Samiti

Prof. Shashi Prabha Kumar
Chairperson
Indian Institute of Advanced
Studies, Shimla

Prof. Shrinivasa Varakhedi
Vice-Chancellor
Central Sanskrit University,
New Delhi

Prof. Joy Sen
Professor
IIT Kharagpur

Prof. Girish Nath Jha
Professor, SSIS Jawaharlal
Nehru University

Prof. Shantanu Chaudhary
Former Director, IIT Jodhpur

Prof. M. D. Srinivas
Chairman
Centre for Policy Studies,
Chennai

Dr. Jitendra B. Shah
Director, Shrut Ratnakar,
Former Director,
L. D. Institute of Indology,
Ahmedabad

Table of Contents

Editorial Note

Working Group Reports

I. Decipherment of Ancient Scripts: Indus, Gilgit, and Sankha	1-29
II. Survey, Documentation, Metadata Standards, and Digital Archiving	30-113
III. Manuscriptology and Palaeography	114-137
IV. Digitization Tools, Platforms, and Protocols (HTR, AI, IIF)	138-159
V. Conservation and Restoration of Manuscripts	160-170
VI. Decoding Manuscripts: Pathways to the Indian Knowledge System	171-214
VII. Manuscripts as Tools of Cultural Diplomacy	215-225
VIII. Legal and Ethical Frameworks for Manuscript Preservation and Access	225-248

Editorial Note

The Gyan Bharatam has been conceived as a national effort to reclaim, revitalise, and reposition Bharat's manuscript heritage in the contemporary world. What sets Gyan Bhartam apart is both its vision and its method. From the outset, it was recognised that safeguarding manuscripts could not rest on isolated initiatives; it required a structured, collaborative, and sustained engagement at the highest level.

Keeping this vision in view, a three-day International Conference—Reclaiming India's Knowledge Legacy through Manuscript Heritage—is scheduled for 11–13 September 2025 at Vigyan Bhawan, New Delhi, marking the mission's ceremonial launch. Notably, the conference coincides with the 132nd anniversary of Swami Vivekananda's iconic address at the World's Parliament of Religions in Chicago in 1893—an intentional echo of India's global intellectual resonance. For this International Conference, the decision to constitute thematic working groups was initiated at a high-level meeting of manuscript experts, laying the foundation for both depth and inclusivity in the Programme's approach. In alignment with Gyan Bharatam vision, eight thematic working groups were formed, each entrusted with a distinct yet interconnected dimension of manuscript heritage. Together, they brought into conversation eminent voices from scholarship, technology, conservation, law, and policy with the following eight themes:

- Decipherment of Ancient Scripts: Indus, Gilgit, and Shankha
- Survey, Documentation, and Metadata Standards & Digital Archiving
- Manuscriptology and Palaeography
- Digitisation Tools, Platforms, and Protocols (HTR, AI, IIF)
- Conservation and Restoration of Manuscripts
- Decoding Manuscripts: Pathways to the Indian Knowledge Systems
- Manuscripts as Tools of Cultural Diplomacy
- Legal and Ethical Frameworks for Manuscript Preservation and Access

The working groups engaged in rigorous dialogue and collaborative inquiry, and their deliberations were not limited to academic exchanges. These groups critically examined current practices, identified systemic gaps, and charted pathways for reform and innovation. The reports that have emerged are therefore not simply collections of viewpoints, but structured interventions that combine empirical evidence, conceptual clarity, and pragmatic recommendations.

For this International Conference, these reports assume special significance. They not only form the backbone of Gyan Bharatam but also serve as a roadmap for action, guiding how manuscripts can be preserved, digitised, studied, and integrated into the nation's intellectual and cultural life. Their insights will also directly contribute to the drafting of the Delhi Declaration on Manuscript Heritage, a landmark statement of shared intent and collective vision for the future.

These reports represent collective intellectual milestones. They embody the spirit of collaborative scholarship and reflect the seriousness with which this endeavour has been pursued. Importantly, they provide value-added insights: diagnosing present challenges while offering concrete, future-oriented solutions that position manuscripts not only within frameworks of preservation but also within the broader spheres of education, technology, diplomacy, and cultural life.

Through these contributions, the Gyan Bharatam transforms manuscript preservation from a specialised pursuit into a national and global endeavour of pride, accessibility, and innovation. It provides a roadmap that carries the potential to place India at the forefront of global leadership in cultural heritage.

We take this opportunity to express our sincere gratitude to the Hon'ble Minister of Culture & Tourism, Shri Gajendra Singh Shekhawat, and to Shri Vivek Agarwal, Secretary, Ministry of Culture, for their constant support and guidance. We also extend our deep appreciation to the Members of the Organising Committee and Advisory Committee for their immense

contributions, and above all, to all the members of the Expert Groups, whose valuable inputs—delivered within such a short span of time—have laid the foundation of this Mission.

Special thanks are due to the Coordinators of all the Thematic Working Groups for their excellent coordination and leadership in steering the deliberations across diverse domains. We also gratefully acknowledge the contributions of all staff and officials of the Ministry of Culture and IGNCA, whose tireless commitment made it possible to accomplish this near-impossible task within such a limited timeframe.

Finally, we place on record our heartfelt thanks to all the distinguished experts, scholars, researchers and manuscript repository holders who have contributed to making this important exercise possible.

As Editors, we warmly welcome all participants of this International Conference. May the deliberations here—grounded in the wisdom of the working groups and enriched by diverse perspectives—guide the Gyan Bharatam towards shaping a resilient, inclusive, and globally relevant future for India's manuscript heritage.

- Editors



Working Group I

Theme: Decipherment of Ancient Scripts: Indus, Gilgit, and Sankha

Coordinator

Prof. Alok Tripathi

Additional Director General (Archaeology), ASI

Co-coordinator

Smt. Priyanka Chandra

Director, Ministry of Culture, Govt. of India

Experts

Dr. T. J. Alone

Joint Director General, ASI

Shri Hemasagar A. Naik

Director (Excavation), ASI

Dr. Bhuvan Vikrama

Director (PDU-IA), ASI, Greater Noida

Dr. Goutami Bhattacharya

Superintending Archaeologist, PDU-IA, ASI, Greater Noida

Dr. Aprajita Sharma

Assistant Superintending Archaeologist, UAW, ASI

Dr. Deepti Agnihotri

Assistant Superintending Archaeologist, ASI

Decipherment of the Indus Script (Current Status and Way Forward)

Introduction – Harappan Civilisation

In hindsight, it is indeed poetic that the earliest evidence of the lost civilization that came to light should remain the enduring enigma even more than a century later.

The abandoned cities, towns, and villages of the Harappan civilization, lost to the world due to the vagaries of nature, continued to remain lost, though their land was trampled by marauding troops and scanned by the discerning eye of an antiquarian. The first mention of these lost cities is perhaps found in the narration of Aristoboulos, an emissary of Alexander the Great, in 326 BCE. He mentions “an abandoned country, with more than a thousand towns and villages deserted after the Indus changed its course”.

The announcement made by John Marshall (1924) of the discovery of the Harappan Civilisation at sites like Harappa and Mohenjo-daro not only pushed back the antiquity of India to the 3rd millennium BCE but also brought it into parity with the other known civilisations of the Old World – Mesopotamia and Egypt. Later years and decades were invested in finding and excavating similar sites, highlighting the similarity of material culture through the finds of artefacts like beads, weights, pottery, lithic tools, bricks, and other objects, as well as in the layout and planning of settlements. The extensive research and fieldwork established the extent of the civilization from Sutkagendor in the west to Alamgirpur in the east, and from Manda in the north to Daimabad in the south, making it the largest in terms of spatial extent.

Of the many objects displaying uniformity and some sort of standardisation, seals stand out for superior craftsmanship as well as for the mystic engravings on them of some kind of meaningful message that has, so far, eluded decipherment.

Origin and Evolution of Scripts

All the scripts of the world originated from the pictorial stage. In the initial stage, the figures used in writing would denote an actual object of similar graphic features. The picture-word would thus represent a tangible thing in written form. In the second evolutionary stage, certain meanings were alluded to the pictures that did not necessarily match the actual pictorial identification. Sometimes, this second meaning could have a close relation with the picture or be completely unrelated. But the new meaning assigned will be agreed upon by the users. In many cases, the new meaning might be an intangible concept that cannot be expressed as a picture.

The repertoire of symbols increased by modifying existing ones or introducing new ones and assigning an agreed-upon meaning to them. Pictographs and abstract signs both existed in the corpus of writing, though the former would dominate. Over time, the primary meaning of certain symbols may evolve. Some symbols would die out, being infrequently used, while new ones would take their place. However, by this stage, some rudimentary guidelines would be in existence for using symbols to convey a message structure.

An important evolutionary stage in the development of script was assigning phonetic values to the picture-words. This gave rise to a range of words that sounded the same (homophones) but had widely differing meanings. Therefore, it was necessary to understand the context in order to identify the appropriate form of the word intended for communication. Subsequently, this ambiguity was sought to be addressed by the use of additives to help in 'semantic determinations. Additional meaning could be affixed to signs by the addition of ligatures – when more than one symbols are combined to form a new symbol.

All these progressions, necessitated by the desire to communicate and/ or document more complex ideas and messages, led to further development in the nature of scripts. All these developments went through a long gestation period. The earliest form of communication thus was restricted to encoding broad ideas and had no internal structure that reflected the spoken language in all its grammatical nuances.

From the pictographic to the logo-syllabic stage, where phonetics came into play more and more. This led to the diminishing of the pictographic features of the motifs, and the signs became easier to form (latter part of the 3rd millennium BCE). Unanimous phonetic values could be assigned to symbols, and words with several sounds could be written efficiently and economically. The next stage in the development was of alphabetic scripts (2nd millennium BCE).

The assigning of syllabic phonetic value to a derived and simplified sign led to the development of the first alphabet. The basic principle of an alphabetic script is the assigning of a symbol for each phoneme (vowel/consonant). This allows for a large number of words with varying pronunciations to be coded efficiently. The phonetic values that were chosen to be represented were usually the initial consonants of the words. In certain languages, vowel sounds were also assigned a sign. The signs represented a solitary vowel sound or a combination with one and/or two consonant sounds.

In contrast, certain texts continued to use pictograms and expanded their repertoire as needed to convey complex ideas. This led to the development of logo-graphic scripts.

The number of symbols used in a hitherto undeciphered script generally provides an initial idea about its nature. The least number of symbols characterizes the alphabetic script. These are usually in the range of 20-35 individual motifs, each representing a phonetic value.

Next comes the syllabic script, which has a range of 100 to 150 symbols. However, the number of syllables needed to write a single word would be very less. A word having a large number of syllables would be present but not used in common parlance.

The largest number of symbols is required in logo-syllabic and logo-graphic scripts – comparatively more in the latter. They may generally range in the neighbourhood of 400-700 characters. Here again, some characters would be more frequently used, while others would be rarely used. Here, the number of signs typically depends on two factors. Firstly, whether the script as whole tends to be leaning more towards the graphic or the phonetic side. The former case would necessitate more characters. Secondly, in the case of phonetic characters, whether they are monosyllabic or polysyllabic.

Indus Script

The discovery of the ‘most curious object’ at Harappa, in 1872-3, brought to light a hitherto unknown script engraved on a ‘stone seal’. Since then, this script has successfully defied any accepted decipherment and has remained the most fascinating and enigmatic chapter in India’s archaeological annals. The first description of this script by an archaeologist read, “Above the bull there is an inscription in six characters, which are quite unknown to me”. For over a century and a half since that discovery, the above description still stands, and the script remains an enduring mystery that has attracted the attention of scholars and amateurs alike.

The discovery of 1872-73 was not a solitary phenomenon for long; discoveries of similar engraved seals, again from Harappa, were reported in 1886 (Dames) and 1912 (Fleet). The archaeological terra firma for these isolated finds was provided by excavations at Harappa and Mohenjo-daro during the 1920s. A large number of findings of these seals from the excavations paved the way for more streamlined research on the objects themselves, as well as on the script and figures they preserved.

Characteristics

With the increase in the number of excavated sites of this culture, many more pieces of evidence of this script, in addition to the well-known seals, came to light. The script was found as seal impressions on terracotta, pottery; steatite, terracotta

and faience tablets; copper and bronze tablets, tools and weapons; ‘miniatures’ of terracotta, stone, faience; small bone or ivory sticks, terracotta bangles, cones, a drainpipe, shell ladle, ivory plaque and die, stone balls and a broken slab, and a bull figurine. However, undoubtedly, the largest collection of the Indus script is on seals. They typically follow a uniform pattern, featuring a line of engraved characters at the top and a pictographic depiction below.

Scholars have posited a relationship between the text and the iconography of the seals. In several cases, the same arrangement of the Indus script has been recorded with different motifs.

Certain characteristics of the Indus script have been found to have comparatively wide acceptance, if not universal. One such feature is regarding the direction of its writing. Most scholars are united in accepting that the script was written from right to left (Gadd, 1931; Hunter, 1934; Ross, 1938; Lal, 1966; Zvelebil, 1970; Alekseev, 1976; Mahadevan, 1977; Parpola, 1994). In the seals, which have a negative rendition of the texts, the reading is from left to right. In contrast, in seal impressions or sealings (positive), the reading is from right to left. This is suggested by the utilization of space, especially on seals. In the case of short texts, the text would start from the right edge, leaving unused space on the left side. Similarly, in the case of longer texts, overlapping and cramming of signs are noted as the text progressively moves to the left, resulting in space constraints. The exception here is when the text is continued in the second incomplete line – it carries onward from left to right. This then gives rise to the boustrophedon. Nine irrefutable pieces of evidence of boustrophedon writing are cited in the complete corpus (Robinson, 2002). However, if the text spans two or three complete lines, each line is written from right to left. According to Mahadevan’s concordance, 6.6% examples are of left-to-right writing.

Possehl identifies the Indus script as a pictographic system. This is the first stage in the origin and development of the writing systems of the Old World. A stage where the sign or the symbol was still the smallest meaningful unit (morpheme), combining in it the root idea with grammatical markers, but the grammatical markers were not reflected in the script and had to be understood. The presence of a substantial number of such standalone signs in the script is likely what led Heras (1953) to propose that the Indus script was a picto-phonographic system.

The Mesopotamian writing system also underwent a similar evolutionary phase, characterized by a pictographic approach. The Proto-Elamite or Linear Elamite script of the Iranian Plateau is also a pictographic script. The geographical proximity of some of the sites of the latter and their near contemporaneity suggests a better correspondence between the Proto-Elamite and the Indus scripts. Hunter (1932) noted the close resemblance of many of the symbols between the two, suggesting

a common/related point of origin. The proximity of the two scripts, though both undeciphered, led Fairservis (1992) to give a transliteration of a Linear Elamite tablet into the Indus script.

The symbols of the Indus script have been identified as belonging to the ‘logo-syllabic’ evolutionary stage of emergence and development of writing (Zvelebil, 1970; Mahadevan, 1988; Parpola, 1994). It is the stage where the symbols not only represented a basic idea or narration – a step up the evolutionary ladder from the pictorial stage – but had developed a rudimentary form of grammatical affixes. The complexity arises here because the grammatical affixes or graphemes could have any number of connotations. Currently, the corpus symbols in the Indus script are divided into two sets: 400-450 (Parpola, 1994; Mahadevan, 1977) and 702 (Wells, 2015).

Having determined the evolutionary stage of the Indus script as logo-syllabic does not suggest a clear break from the antecedent stages altogether. Several symbols in the script still conform to the earlier solitary written word stage – for example, the ‘fish symbol’.

Wells believes that pottery markings found across sites having the same cultural affiliation are not part of an integrated process. In fact, they are strictly local traditions having similarity only in their cultural contexts. The graphic similarities in these markings can be and should be explained as resulting from geographical proximity, trade, and above all, the universal character of certain symbols (‘graphic universals’ as he denotes them), regardless of time and space. He states that the affinity seen by scholars between Indus and Proto-Elamite scripts – both undeciphered – can be explained based on the above. They do not have an inherent common origin, but rather two different autochthonous development models. The presence of Wells’ ‘graphic universals’ in both the scripts may make them seem related where no relations exist.

S.R. Rao is an exception in identifying the script as alphabetic (Rao, 1982). He believes that the Indus script symbols, as found in the archaeological record of the Mature Harappan period, are 62 basic signs. The numbers were further reduced to 20 during the Late Harappan period.

The Indus symbols denote both meaning as well as the phonetic sound that goes with it. The symbol, as it appears in the text, may stand alone for either the meaning or the sound, or for both.

The number of Indus symbols is an equally significant database for the successful decipherment of the Indus script. It not only identifies and fixes the signs that have to be deciphered but also identifies variants that may have a similar connotation in the system.

There is no unanimity among scholars regarding the number of Indus symbols. The primary point of contention among scholars is whether to categorize the symbols as basic symbols or as variant forms of basic symbols. The first listing by Langdon in 1931 reported a number of 288. Hunter's list of 1932 lowered the symbols to 149. The subsequent lists all gave out an increased number of symbols: 270 - von Meriggi (1934), 537 - Dani (1963), 417 - Mahadevan (1977), 396 - Koskeniemi and Parpola (1982), 419 - Fairservis Jr. (1992), 694 - Wells (2015), 713 - Fuls (2023).

Another useful dataset is the frequency of each symbol and its associated variants. In Mahadevan's concordance, 113 symbols occur only once; 47 only twice, and 59 signs have fewer than 5 re-occurrences. Only 20 symbols have over 50% occurrences, 8 symbols have 31% and only 4 have 21%. In Wells' list, nearly 50% symbols occur only once.

The Indus script also contains numerals or numeral-like signs. The strokes found in association with the basic signs denote numerical value (Waddell, 1925; Ross, 1938; Kinnier Wilson, 1974; Mitchner, 1978; Newberry, 1980; Mahadevan, 1988; Fairservis, 1992). The issue with strokes is that there are clearly two types: short and long ones. The short ones have been found in combinations of 1-10 and 12; interestingly, 11 is absent. The longer ones are found in groups of 1-7. Another symbol that has been attributed a numerical value is the semi-circle – for tens (Mahadevan, 1988; Parpola, 1994).

The prevalence of the numerical system in the Harappan culture was noted by Marshall in connection with the weights discovered during excavations. He proposed a mixed numerical system based upon binary, octonary, and decimal (Marshall, 1931). Ross posited a twin numerical system based on ten and eight (Ross, 1938). Similarly, Fairservis favoured a system with a base of eight, while Mahadevan advocated for a decimal base (Mahadevan, 1988; Fairservis, 1992). According to Fairservis, the fact that there are no groupings of longer strokes beyond 7 is suggestive of a numerical system based on 8. He was further convinced by the evidence for a Dravidian base-8 system.

Seals bearing an animal depiction invariably bear a text. This suggests a close relationship between the iconographic depiction and the text. The idea suggested by the picture is complemented and supplemented by the text, and both are integral to a complete understanding of what is being conveyed.

The brevity of the Indus script is yet another near-universal feature. That more than half of the recorded examples of the script come from the seals makes the above statement a logical one. The average length of the text is five symbols, with the longest being of 26 symbols inscribed on three sides of a triangular terracotta prism.

There can be several plausible explanations (both external and internal) for the short text length of the Indus script. Firstly (externally), the information or message recorded on them was probably an identification of the owner – personal names, epithets, protective invocations, formulaic dedications, benedictions, etc. Secondly (internal), the preponderance of logograms in comparison to syllabic characters. Finally (internal), the use of ligatures (sign combination representing a single sound or unit of meaning), which suggest agglutination (process of creating new words encoding a phrase by adding morphemes), leading to brevity.

There is uniformity in the Indus script. There is geographical homogeneity in the usage of signs across the length and breadth of the land.

Evolution of the Indus Script

The presence of a potter's mark, both pre- and post-firing, is considered the embryonic stage in the origin and evolution of script in a pre-literate society. These rudimentary symbols may or may not survive the progression and form part, partially or wholly, of the actual script as and when it develops. The Early Harappan potter's marks were mainly strokes, in groups, which would attest to their numerical connotations. Archaeological evidence of potter's marks has been profusely recorded from nearly all excavated sites for several centuries prior to the appearance of the Indus script. In a stratigraphic context, at several avowedly Harappan sites, there is a continuous sequence of findings of potter's mark from the Early to the Mature Harappan period.

Despite the above, scholars cite the lack of evolutionary stages in the different symbols that make up the Indus script during the Mature Harappan period, as well as the Early Harappan period, as a glaring gap (Parpola, 1994). The lack of internal evolutionary stages of script progression is cited as evidence for external influence in the development of the Indus script, as it appears during the Mature Harappan period. The Proto-Elamite script, with its vague similarity to certain elements and geographical proximity, is the closest contender; however, it is unconnected in time, being a few centuries older than the Mature Harappan evidence.

The establishment of sea trade relations with the Gulf during the middle of the 3rd millennium BCE is identified as the catalyst in the development of the script during the Mature Harappan period. A 'Late Proto-Elamite' inspiration for the development of the Indus script has been proposed as a better alternative to its early counterpart and is supported by the influence of Proto-Elamite glyptic art on the iconographic motifs depicted on the Harappan seals.

The standardized form of the recorded symbols that appears on seals and other materials during the Mature Harappan period is taken as suggestive of external influence, rather than a generic internal evolutionary process (Possehl 1990, 1996; Parpola, 1994). Moreover, even during its approximately 600-year existence, there is no noticeable change in the form of the symbols, barring those necessitated by space constraints or dictated by the media on which they were engraved.

Given the above evidence, it appears logical that the Harappan traders, in the course of their dealings with their western counterparts, were exposed to the novelty of record-keeping and communicating contractual obligations. This external stimulus, though potent, was restricted to just that – the idea of recording and/or communicating. The idea, when put into action in the Indus plains, gave rise to a script that was foreign in idea but indigenous in its execution. Thus, the symbols do not exhibit any external influence and were entirely native creations.

Contrary to the above, several scholars have disputed the external influence on the development of the Indus script. As early as 1931, Gadd categorically denied any influence from Sumeria or, for that matter, even Egypt on the Indus script (Gadd, 1931). A tentative correlation between the graffiti found on the pottery of several pre-Harappan cultures with the later developed Indus script was suggested by Fairervis (Fairervis, 1959).

B.B. Lal was one of the early pioneers in the study of the indigenous origin of the script and its continuation into Late Harappan and other extra-Harappan chalcolithic cultures. Similarly, S.R. Rao also favours a long gestation period for the Indus script from the Early Harappan period (c. 3300 BCE), its most flourishing phase during the Mature Harappan period, to its disappearance during the Late Harappan period. He identifies the script as logo-syllabic only during its early phase, quickly transforming into alphabetic during the Mature Harappan period. The indigenous origin of the Indus script in the river valleys of the Indus and Ghaggar-Hakra, as well as in Baluchistan, between the 4th and the first half of the 3rd millennium BCE, coeval with the Early Harappan culture, poses a credible challenge to the external inspiration theory of earlier times.

Based on the findings from recent excavations and previously known data, Kenoyer (2020) posits an internal, stage-wise progression in the origin and development of the Indus script. The first stage is related to evidence from pre-firing potter's marks and post-firing graffiti. This stage is identified as the proto-script stage, a uniform phenomenon observed across all Old-World pre-literate cultures. Some of the symbols recorded on pottery have continued to exist even after the script developed and gained widespread use.

The second evolutionary stage, referred to as the Early Indus script by Kenoyer, is identified by evidence of pre-firing inscriptions, including the use of graphic symbols. New symbols, one to three symbols that form an integral part of the developed Harappan script, appear as post-firing graffiti. The first evidence of inscribed seals, featuring one or two symbols, also dates back to this stage. This stage corresponds to the final phase of the Early Harappan cultural period.

The third phase is the high watermark age of the Indus script, which corresponds to the Mature Harappan period. Examples of script burst forth in a variety of media, of which the steatite seals remain the major group.

Decline

B.B. Lal shed light on the legacy of the Indus script in the form of graffiti on Chalcolithic pottery from the Deccan, as well as on those of the Megalithic cultures, which extend further south, indicating an Iron Age affiliation. Though the majority of the recorded evidence is of post-firing graffiti, pre-firing incisions are also not unknown in both Chalcolithic and Megalithic contexts. These examples are not confined to any particular type of pottery in either cultural context. In the case of megalithic cultures, pottery with similar graffiti has been found in both burials and habitation sites. Furthermore, no partiality to one motif is noted in burials. In some cases, more than one motif was found in one burial. The motifs on the pots may occur in isolation or in repetition of the same character, and may also bear two or three different motifs together.

In the context of the Deccan, the occurrence of graffiti suggestive of Harappan connection is recorded from a large geographical area and in quite good numbers to be discarded as isolated finds. Chronologically, they are placed in the 2nd millennium BCE. The similarity of some pottery types between sites in Gujarat and those of the Deccan (first half of the 2nd millennium BCE) suggests a connection that can be attributed to the seasonal transhumance of pastoral nomads between the two regions, which also continued until very recently. This then suggests a plausible reason for the continuance of Indus script graffiti in the Deccan.

Parpola opines that after the end of the Mature Harappan period Indus script shows a sharp decline. The rudimentary lingering did continue at some sites, but it was the last remnants that persisted for a long time (Parpola, 1994). He dismisses the 'Late Harappan linear script' theory based on incisions on pottery from Rangpur and other post-Urban sites in Gujarat outright. Although he accepts the isolated presence of some Indus symbols, they were merely standalone symbols that did not carry forward the legacy of the Indus script.

However, Kenoyer identifies this final phase of the Indus script as coeval with the Late Harappan period. Examples of graffiti on pottery are cited as examples of the final stage of the Indus script.

Is it a Script?

The lone discordant voice in the entire gamut of literature on the Indus script is from Farmer, Sproat, and Witzel (2004; Farmer, 2006), who defiantly highlighted the ‘Myth of a literate Harappan Civilization’. According to them, the brevity of the inscriptions proves that the Indus symbols were not meant to record long documentation of any sort, and therefore, not a script. The symbols on different media do not encode any language. Other arguments that they give in favour of their proposition are the frequency and repetition patterns of symbols, and also the fact that a large number of them occur only once.

They argue that there is no missing manuscript that has to be found or that could have been written down on perishable materials. What is known to date should be taken to represent the entire repertoire of Indus symbols. Another argument put forward by Farmer is the absence of any sort of evolution in the morphology of the motifs, which should be a distinct characteristic had it been a legitimate script owing to frequent scribal usage.

They propose that once the entire script myth is debunked, serious research can be conducted on the nuances of how ‘the biggest non-literate civilization’ functioned and operated on such a large scale across space and time. The symbols saved from the burden of being a script can be studied for their ‘untapped data’ and the meaning of their variations and regional and/ or sub-regional antecedence, if any.

They believe that scholars have hitherto overemphasized the trade and accounting aspects in interpreting the Indus symbols. To them, most of the symbols are suggestive of farming/ agriculture, and some of them could be political-religious symbols representative of ruling classes, administrative offices, clans, or seasonal agricultural rites and rituals specific to micro eco-zones.

Farmer et al. also fail to acknowledge the utility of statistical analysis based on corpora and concordances. They argue that the ambivalent stage of script development, as witnessed in the 3rd millennium BCE, is not optimum for statistical analysis. Their characteristic internal features, such as preponderance of ‘polyvalent signs’ and omission of ‘syntactic, semantic, and phonetic data’, do not lend themselves to any form of statistical analysis.

Quite expectedly, this proposition received severe and widespread condemnation

from established scholars working in the field. Vidale countered them with their own arguments. He stated that the... Regarding the brevity of symbols as evidence of non-literacy, Parpola points out that "... we cannot expect complete sentences in seals and other types of objects preserved. But even written noun phrases qualify as language-based script" (Parpola, 2008).

If not for anything else, Farmer et al.'s views have stirred the proverbial hornet's nest. They threw in a potent out-of-the-box idea into an already heady mix of innumerable claims and counterclaims.

Language

One of the several challenges that face scholars with respect to the Indus script and its plausible course of decipherment is the identification of the language. The two rival claimants to this question had a simultaneous origin in 1931 in the same volume. While Langdon offered the Indo-Aryan language family as the answer, Marshall lent his weight to the Dravidian family group.

The geographical extent of the known excavated and explored sites suggests that several languages may have been spoken in their various regional pockets. However, the characters of the recorded Indus script examples from sites located in different geographical regions, along with consistent structural patterns, point towards an unambiguous uniformity suggestive of one widely accepted language, at least for writing purposes. This premise is further strengthened by the evidence of the Indus script found on seals from sites in Mesopotamia and the Gulf. The internal evidence from the seals suggests that the script was used in these sites to convey messages in a different language. This is inferred on the basis of distinctions noted in the sequencing of symbols from those usually found in the sites of the Indus plain.

The geographical location of the Indus sites and the positioning of this geographical region in relation to the languages prevalent in neighbouring regions provide the first tentative step towards identifying the language family, if not the language itself, of the Harappans. Based on the above, a fairly wide space, encompassing Indo-Iranian borderlands, Central Asia, and, of course, the Indian subcontinent, falls within the ambit where the search for the Harappan language can be concentrated, based on contemporary known language families.

Important archaeological evidence that comes to aid is a cylindrical Akkadian seal that has been read as 'Su-ilisu, Meluhha interpreter'. Meluhha, denoting even partially the Harappan people or region, is mentioned for the first time in the cuneiform documents of Mesopotamia during the Akkadian period. Coming back to the inscribed seal under reference, the mention of 'Meluhha interpreter' is significant

in suggesting that the language of the 'Meluhhans' was so distinct from the known languages of Mesopotamia that it necessitated an interpreter. While this does not directly aid in the identification of the linguistic affiliation of Harappans, it helps in narrowing the plausible candidates by eliminating the Sumerian, Akkadian, and West Semitic language families.

An important observation regarding seals and their language found in Mesopotamia is attributed to Hunter (1932). He compares the sequencing of symbols found on the square and round seals there. The sequencing on four-square seals is similar to that found at Mohenjo-daro (or elsewhere in the Indus plains), while that on the round seals is completely new. This difference leads Hunter to suggest that the square seals were imported into Mesopotamia and therefore conformed to the prevailing Indus language, while the round seals catered to the language of 'Sumerian or Semitic-speaking' persons of Indus origin, and hence a change in the sequencing of symbols. This hypothesis is also favoured by Parpola (1994) and is strengthened by the newer discoveries of such marks in the Near Eastern Harappan seal discoveries.

In the geographical region of India, the challenge is of a completely contrasting nature to that of the Iranian Plateau. There is absolutely no undisputable archaeological or literary evidence of the linguistic affiliation of the different cultures that existed in this area during the 3rd-2nd millennium BCE. In present times, the Indian subcontinent is home to three major language groups – the Indo-Aryan, the Dravidian, and the Austro-Asiatic.

The Indo-Aryan language is a sub-branch of the bigger language group, the Indo-Iranian branch, which in turn belongs to the even larger Indo-European family. The Indo-Iranian language further branched out into two distinct linguistic subfamilies – Iranian and Indo-Aryan. Iranian languages were spoken along the north-western borderlands of the Indian subcontinent roughly from the 1st millennium BCE.

The earliest literary records of the region date back to approximately the second half of the 2nd millennium BCE and are written in the Indo-Aryan language. The earliest hymns of the Rig Veda belong to this phase. The linguistic affiliation of these verses is a prototype of Classical Sanskrit from later periods. Panini's grammar of the second half of the 1st millennium BCE is based on this Classical Sanskrit language. The geographical space that is identified as concomitant with the composition of the Rigvedic hymns is the Swat valley and the Punjab plains.

S.R. Rao (1982) identifies the underlying language of the Indus script as 'Indo-Aryan' – Sanskrit. He arrived at this conclusion from a circuitous route that took him to assigning Semitic phonetic values to Harappan symbols. The earliest evidence of the Semitic script dates back to around 1500 BCE. Rao found similarity (as high as 75%) between the linear Semitic script and the Late Harappan linear symbols.

Reason being, according to him, that the Semitic people modelled their script on the latter. This led him to replicate the Semitic phonetic values in the corresponding Late Harappan motifs, which ultimately led to the discovery of the language as Indo-Aryan.

The advocates of the Dravidian model of language base propose that the Dravidians entered into the subcontinent from the northwest in the 4th millennium BCE (Parpola & Parpola, 1975; Parpola, 1994; Zvelebil, 1972, 1990; Thapar, 1975; McAlpin, 1981). Thereafter, a first break occurred within the Proto-Dravidian language family, and this initial breakaway group was designated as the Proto-North Dravidian. This group is currently represented by the Brahui language, which is spoken by a small community living in the Brahui Hills of southern Balochistan. Other extant languages of this Proto-North Dravidian family are Kurukh and Malto, spoken by groups in parts of eastern central India.

The Proto-South Dravidian further bifurcated into Central and South Dravidian sometime around 1500 BCE. Central Dravidian is a minuscule linguistic family centered around small groups in Maharashtra and Odisha speaking Kolami, Naiki, Parji, and Gadaba. The South Dravidian further split into a southern branch comprising Kannada, Malayalam, Tamil, and a number of languages spoken by very few people. The south-central branch comprises mainly Telugu speakers, with small pockets in Maharashtra, Madhya Pradesh, and Odisha, who also speak the Gondi language.

Mahadevan based his decipherment of the Indus script on the identification of the language as belonging to the Dravidian linguistic family (1972). In support of his claims, he presents evidence, including a definite identification of a Dravidian linguistic substratum within the Vedic literature. This has been taken by Mahadevan as proof of the existence of a Dravidian-speaking populace in the northwest. Based on statistical analysis, Mahadevan asserts that the Indus script appears to be more suited for the Dravidian family of languages than for the Indo-European linguistic family.

An interesting hypothesis was put forward by McAlpin regarding the languages spoken in the 3rd and 2nd millennia BCE. He gave the concept of a 'Proto-Elamo-Dravidian' language family covering a large geographical area encompassing south-eastern Iran, including the plateau region, Pakistan, and northern India (McAlpin, 1981). This language family, an admixture, and not Dravidian itself, which was used by the people of the Great Indus plain and formed the base of their script.

The Austro-Asiatic language speakers in the country are very less compared to the Indo-Aryan and Dravidian language families. Until two centuries ago, the Austro-Asiatic language was mostly unwritten, consisting of spoken words. The linguistic sub-families associated

with these groups are restricted to tribal populations in the country, with spatial distributions limited to the eastern frontier and a few pockets in central India.

A new underlying language hypothesis was forwarded in 1998 by Witzel. He proposed Proto-Munda as the underlying language of the Indus script. The basis for this was the practice of prefixing words of non-Indo-Aryan origin in Vedic Sanskrit. This structural peculiarity suggested to him an Austro-Asiatic language. A practice that was also prevalent in the Munda language. He identifies the language as Para-Munda. This language would then be the western-most branch of the Austro-Asiatic linguistic family. The geographical scope of the language is identified as extending to eastern Punjab, Haryana, and further eastwards.

Witzel, however, identifies the main underlying language as Proto-Munda; he admits that the primary spoken language during the second half of the 2nd millennium BCE was likely a dialect – Meluhha.

Methodology

One of the easiest and most self-explanatory ways to start deciphering is with the pictorial motifs themselves – stick figures, fish, and so on. These could be suggestive of the words associated with the motif and homonyms – different unrelated meanings but the same word. This method relies heavily on the accurate identification of the underlying linguistic affiliation of the script to distinguish the word from its homonyms.

In the case of logo-graphic script, the motif may have undergone alterations in its meaning over time and across different contexts. Polyphony and polysemy are other characteristics of logo-graphic scripts. As a result of which it may be difficult, even in deciphered scripts, to affirmatively identify the correct value of the symbol in a given context. Alternatively, in a logo-syllabic system, the same may have changed from the original morpheme, retaining only its phonetic value.

A script that has a long lifespan gradually evolves internally. The basic symbol is added upon through time – grammatical or semantic additives, diacritical marks, phonetic affixes, semantic determinatives, etc. Sometimes the symbols themselves undergo modifications, giving rise to new symbols.

Another methodology that has been used with mixed results and varying reception is to select and choose symbols that can be interpreted pictorially, obviously, and may also be phonetically, from a candidate linguistic family. Homonyms from the identified language are also introduced into the playing field, allowing for varying interpretations or meanings. The basic idea is to find as many symbols that can be assigned meaning or phonetic values in the chosen language, then find a sequence

of such symbols that convey a meaningful message/idea through words, phrases, and sentences again in the same language. However, when the same meaning with the same combination of symbols is attempted in any other candidate language, the result is non-coherence.

The above is exactly what has been attempted for Indus symbols with respect to various candidate languages, including Dravidian (proto-Dravidian) and Sanskrit. However, there are obvious flaws with this approach; not the least of which is the wide leeway accorded for interpretation. With such a long rope available for interpretation, scholars and experts, often influenced by biases, have overinterpreted possible meanings and arbitrarily stretched the limits too far.

Pictorial Bilinguals

The pictorial interpretation of some Indus symbols, in a few cases, is supported by the iconography revealed in Harappan art forms, as depicted on seals and elsewhere. The commonly found ‘jar symbol’ is one such motif. In many tablets, the iconic representation shows a half-kneeling or sitting figure offering a ‘jar’ to a tree or a seated figure. In this instance, the tree or figure appears to suggest a divine entity being propitiated. This graphic representation and its identification support *ipso facto* the interpretation of the ‘jar’ motif as actually a jar containing an offering and hence its cultic/religious connotation.

The copper tablets from Mohenjo-daro also represent ‘pictorial bilinguals. They have small texts on one side and an image on the reverse – an animal figure. The animal is considered a representation of the deity. Sometimes, the reverse does away with the figurative depiction and instead has a short text. Parpola (1994) believes that the inscription that replaces the figure is the name of the deity.

Rebus Principle

As mentioned earlier, the development of the script started with pictorial depictions. The pictures were easily identifiable with objects that they denoted in the language known to the specific group of people to which they catered. In the second stage, the same picture could denote any other word with the same phonetic values but different meanings. Such words were known as homophones. Therefore, in the second stage of script development, the word-picture was used to denote an abstract concept that sounded the same but could not be depicted by pictures, such as nouns, abstract concepts, and proper names. This gave rise to logo-syllabic scripts, and the methodology employed was known as ‘rebus’ writing. ‘Rebus’ is a Latin word meaning ‘by means of things’

Markov Model

The Markov model is a mathematical process that describes a system transitioning between states over time. The key is that the probability of transitioning from one state to another depends on the current state, not on different antecedent states leading up to it. The same model has been applied in attempts to decipher the Indus script (Rao et al., 2009). The model has been used to identify the statistical and syntactic features of the script.

The symbols of the Indus script are treated as ‘states’ within the Markov model and identify the probability of one symbol following the other in a sequence. It has identified some symbols as most likely to be found at the beginning or end of a text. According to this model, it has been concluded that the most frequently occurring symbol in the entire corpus is less likely to come at the beginning but most likely to be at the end. The statistical properties, as revealed by the application of the Markov model, strongly suggest that the script was well-suited for communicating a structured language. The model, when applied exclusively to the Indus seals found in Mesopotamia, revealed changes in the structure of the symbols, suggesting that the content differed widely from what was prevalent in the Greater Indus plains. Certain combinations of Indus symbols were noted on the seals from foreign shores that never occur in the Indus region.

Another area where the Markov model is beneficial is in predicting missing or illegible symbols. The efficacy of the Markov model in correctly identifying the missing or illegible Indus symbols was verified by testing the model with deliberately hiding symbols for test purposes. The results gave a 75% accuracy.

The Markov model also sheds interesting light on the linguistic nature of the Indus script. The linguistic nature is determined by the degree of entropy and is compared with that of known languages. In non-linguistic systems, entropy is at the higher end, suggesting that the symbols do not conform to any formal system and are used randomly. In contrast, linguistic systems exhibit varying degrees of order in the sequences of words and characters, ranging from rigid to random. The entropy of the Indus script, when analyzed using the Markov model, indicated that it fell within the range of natural languages.

Attempts

Cunningham, who reported the first seal of this ancient civilisation, came up with perhaps the first attempted identification of the script. He identified it as a hitherto unknown prototype of the Brahmi script of the historical period.

After Marshall’s announcement of the discovery of the Indus civilisation in

September 1924, Gadd and Smith (October 1924) presented their first attempt at identifying the similarities between the Indus symbols and known Sumerian motifs. Similarly, Sayce (September 1924) also identified similarities between seals from Harappan sites and those from Mesopotamia and southwest Iran.

In 1925, Waddell, based on the photographs of seals published in Marshall's news report of September 1924 and those from Harappa (5), totalling 19 in all, published the first book on decipherment of the Indus script. He attempted to decipher the script based on its apparent Sumerian association. His association stemmed from the belief that the Sumerians and the people inhabiting the great Indus plains were racially related to the Aryans and spoke a language belonging to the Indo-European linguistic family. His approach was to extrapolate the phonetic values of Sumerian signs to similar-looking Indus symbols. On this basis, he read the names of Sumerian kings and Aryan kings mentioned in the Vedas and in the Sanskrit epics.

Barton published a few reports on Indus inscriptions known at the time (1928, 1929, 1930). He highlighted that some of the motifs had wide-ranging parallels in other writing systems as well – Sumerian, Elamite (Proto-Elamite), Hittite, Egyptian, Cretan, Cypriot, and even Chinese. He also opined that the Harappans' number system was a decimal system, and that the larger strokes represented counts of 10s. He did not claim to have deciphered the script, but he still provided interpretations of some texts, accompanied by a caveat of caution.

Waddell also gave his unconditional support in identifying the underlying linguistic base of the script as Vedic and Epic Sanskrit – 'the only available authentic key'. Some motifs are common to both scripts, as has been subsequently reiterated by several scholars. However, some of Waddell's interpretations are way off the reasonable bandwidth of plausibility. In seals, where texts have added pictorial depictions, he assigns phonetic values to the heads of the animals depicted, such as unicorns and zebras. Limiting the association of the pictorial depiction to only phonetic value in interpreting the inscribed text is a myopic view.

Gadd, Smith, and Langdon collaborated on the report of Mohenjo-daro and credibly expanded on the information from the latest discoveries. Gadd opined that the Indus symbols were pictorial representations of real-life objects – such as a man in various attitudes, a fish, a bird, an animal, an arrow/spear, among others. He identifies the total repertoire of symbols as being around 300, but provides definite identifications for only 25 of them.

Langdon, who also contributed a chapter on the Indus script in the report on Mohenjo-daro (1931), asserts that the script formed the basis for the Brahmi script of the historical period. In favour of the hypothesis, he presents a comparison chart

of Brahmi and hitherto discovered Indus symbols. He also identifies some motifs based on their iconography. Langdon was also the first to suggest that the direction of scribing was from right to left. He also presents a sign list which had 288 Indus symbols along with their variants.

Scholars such as Sayce, Gadd, and Smith have highlighted the resemblance between the Indus script and Proto-Elamite pictograms, as well as archaic Sumerian scripts.

Pran Nath (1931, 1932) was one of the early Indians who invested regularly in the puzzle of the Indus script. In line with contemporary fashion, he also believed that the Brahmi script evolved from the above and was an alphabetical script. He attempted to identify some of the symbols found on the punch-marked coins with those from the Indus script. With his self-assuredness regarding Brahmi, he went on to transliterate 78 Indus texts. This led him to identify the names of Aryan and Sumerian deities.

An eminent archaeologist who succumbed to the lure of decipherment was Sir Flinders Petrie (1932), the famed Egyptologist. Biased by his experience with Egyptian scripts, he assumed the Indus script was also pictorial. He identified the seals had official connotations and the text contain titles, designations, etc. For example, he identified the jar sign as ‘wakil’ or an agent. He made no suggestion on the plausibility of the language of the script.

A far-fetched association, separated both in space and time, was propounded by Vilmos Hevesy (1932). He compared the Indus script with the rongorongo tablets of the Easter Islands – also undeciphered.

Another comparison with another undeciphered script was attempted by Piccoli (1933). He found similarity of motifs – geometric ones – between the Indus script and the Etruscan.

Hunter’s book of 1934 was based on his doctoral research. In 1932, he published an article regarded as the earliest detailed study of the script based on Marshall’s report. He supported Langdon in identifying the Indus script as a precursor to the historic Brahmi script. He identified the language as Dravidian – a form of Proto-Brahui. Hunter also found a close connection between the script and both the Egyptian and Proto-Elamite languages. He presented a chart comparing the symbols of the Indus script with those of the Egyptian, Proto-Elamite, Sumerian, Brahmi, South Semitic, Phoenician, and Cypriot scripts. Hunter did not claim or attempt to decipher the script. His attempts were more in the nature of understanding the inherent nuances of the script itself.

In 1934, Piero von Meriggi attempted to understand the Indus symbols in their pictorial form, rather than attempting to decipher them individually. He also favoured the Indus language with the Dravidian family. He also conducts an internal analysis of the texts. The short vertical strokes are identified by him as word dividers. He also assigns grammatical identification to some signs – the jar sign for genitive case (suggestive of relation/possession); the arrowhead or lance for nominative case (subject of the sentence); the trident for dative ending (indirect object or recipient of an action).

Hrozny's attempts at deciphering the Indus script were based on his successful attempts with respect to the Hittite script. He begins with the familiar, identifying the symbols that have similarity with Hittite hieroglyphs. In the second stage, he replicated the common symbols with the phonetic values from the Hittite language, Indo-European linguistic family, but not Sanskrit. Since the Hittites were a millennium younger than the Harappan civilisation, by inference, the latter script is identified as the progenitor of the former.

In 1953, Father Henry Heras, in his book, provided a detailed narrative and explanation of his hypothesis: that a proto-Dravidian language served as the basis for deciphering the script. He read many of the symbols rooted in their meaning and phonetic sounds from the Dravidian language family – more specifically, the Old Tamil language. He identified the script as picto-phonographic. The symbols, according to him, expressed complete words.

Based on internal analysis of the texts, he also concludes that the jar sign is a terminal symbol. He also posits that the texts on the seals are personal names or epithets, and therefore, when the jar motif occurs at the end, it is usually as a genitive ending. He identifies two categories of symbols – the pictographs and the phonetic signs. For identifying the meaning of the latter, he depended on similar motifs in other known scripts and took their meaning to account for Indus symbols. For this, he cast a wide net – Sumerian, Egyptian, Hittite, and even Early Chinese. Then, he identified the root word, which suggested the same meaning in the Dravidian language, and applied it to the Indus symbols.

The 1960s marked a watershed decade in the decipherment of the Indus script with the introduction of computer-based analysis. Two independent teams, working in Russia and Finland, were involved. The Russian team, led by Yuriy Knorozov, announced its decipherment objective in 1965. Their early publications on the subject consisted of a series of articles published between 1968 and 1981. Their work needed translations into English and was accompanied by a critical review by Zide and Zvelebil (1976). In their final report submitted in 1979, they provided meanings to all known Indus symbols. However, neither did they publish any text of

the project nor the computational analysis of the symbols.

The computational analysis was undertaken by the team to understand and isolate the ‘positional-statistical characteristic’ of the Indus symbols. This data helped them to identify ‘blocks’ within the texts. These ‘blocks’ that they identified were a combination of symbols that denote morphemes, phonemes, determinatives, numerals, ‘variables’, and ‘semi-variables’. Another feature of the script enumerated by them is the preponderance of suffixes over prefixes or infixes. Based on the above internal characteristics of the script, the team came to the discovery that the underlying language of the script should come from the Dravidian linguistic family.

The team associated the fish motif with the phonetic “min” or “meen” in Dravidian languages. Identifying homophones of the word and their associated meanings was the next stage in the process. The word min/ meen also means ‘bright’ and/or ‘star’. Therefore, the fish motif was identified as referring to both ‘star’ and/ or ‘bright’ in the Indus script. The fish motif, associated with six strokes, is interpreted as ‘arumin’ – the six stars – the Dravidian name for the constellation Pleiades.

The Finnish team, led by Asko Parpola, published the first computer-generated concordance (a listing of the occurrence of each symbol in the entire corpus) of the Indus texts in 1973. A revised edition was published between 1979 and 1982. They proposed some readings of the symbols using the Dravidian language family as the base. A more detailed discussion on the readings was published in the form of a book by Parpola in 1994. The latter tome lists 386 symbols and 12 more unnumbered ones.

The team changed their views regarding the nature of the Indus script from a combination of logographic and logo-syllabic to purely logographic and then back to a combination of logographic and logo-syllabic. They also proposed that the script was written based on the principle of rebus. They also identify the language of the script as belonging to the Dravidian family and that the script is to be read from right to left in the majority of cases.

With the help of computing, the Finnish team was able to identify ‘words’ by internal structural analysis of the texts. This involved identifying the pairing of symbols, the frequency of symbols, and their positional contexts. From the statistical analysis, they were able to identify ‘inflections’ – motifs that appear at the end of texts and help in denoting the end of ‘words/ phrases. They also laid down a 7-point criterion that allows for the identification of ‘inflection’ motifs in the texts.

The first attempt by an Indian to create a holistic corpus and concordance is by I. Mahadevan (1970, 1972, 1973, 1979, 1981). With the help of his datasets, he

proceeds with decipherment based on what he calls ‘method of parallelisms’. This method begins with a series of premises – the short text seals record personal names while the longer may include epithets, honorifics, occupations, place-names, etc.; after the decline of Indus writing, some of the symbols were reinvented as traditional symbols used in iconography, royal insignia, emblems on tokens, coins, seals, others having religious, heraldic, totemic connotations, etc.; gleaning these generational motifs through ‘arbitrariness of symbolism’, ‘absence of convincing Indo-Aryan etymology for loan-words, the use of ‘myth and folk etymology’ to account for both of the above, and identification of distortions that accompany ideas and meanings as they are transferred from one sociolinguistic context to another.

Mahadevan attempted to decipher the symbols in relation to the Dravidian language family. Firstly, he identified the pictorial idea conveyed by the Indus symbol and then sought the phonetic value for the same with the help of Dravidian etymology. He takes the spoked wheel sign and associates it with the Vedic concept of the ‘solar wheel’ – the Sun-God, also the sovereign power. In the Dravidian language family, he attempts to find the etymology for ‘sun-god-king’, which comes to ‘vec/vey/ve’. In Old Tamil, the same is ‘ve-nt-an’ – paramount sovereign. However, the wheel symbolism is lost, but it clearly denotes ‘sun’ and ‘god’.

Another important contribution of Mahadevan (2009), although one that did not gain widespread attention, was the distinction he proposed between decipherment and interpretation, specifically in relation to the Indus script. While decipherment would involve the actual reading of the texts, interpretation would entail understanding the meanings that the texts intend to convey, even if their linguistic and phonetic values remain a mystery. The latter has a higher probability of success.

Kinnier Wilson (1974) added his scholarly weight in favour of Sumerian relation to the Indus language and attempted decipherment based on the same. He proposed a common geographical region for the origin of the two – India. Although he acknowledges the leaning towards differences between the two, he believes that the similarities suggest a common origin before the eventual branching out. The smaller branch migrated to Sumer and the bigger one to the Indus plains. He identified the text on the seals as having purely economic connotations – ‘accountancy’ as he calls it. He identified the number system as following the decimal system.

He also finds the Sumerian name for the Harappan region – Ba’d-imin, meaning ‘the seven high places’. He devotes considerable attention to animal symbolism on the seals and identifies the zebu (*Bos indicus* - a type of buffalo) as symbolizing the ruling power at Mohenjo-daro city. The fabled unicorn representation, according to him, was the iconographic depiction of ‘Bad-imin’ in its entirety.

The next in the series of decipherment was by Chakravorty (1975). He begins with the now familiar assumption that the seals record personal names. In search of deciphering the names, he fast forwards to the Mahabharata. Ipso facto, he believed that the names recorded on the seals were Indo-European, and there is continuity between the Harappan and Epic period, a wide margin both in space and time.

Mitchener (1978) claimed the decipherment of parts of the script based on the Sumerian association of motifs through similarity. He identifies the linguistic base of the script as Indo-Aryan. His starting point takes us back to Kinnier Wilson and Bad-amin. Mitchener believes that the Rig Vedic sapta sindhavas and the hapta hindu of the Avesta are an echo of the 'seven' also found in the Sumerian Bad-amin or 'the seven high places'. This phrase was used in Sumerian to denote a place east of Sumeria and beyond the city of Aratta.

Mitchner's idea was to trace similar connotations in the Indus script that would suggest the Proto-Indians and Proto-Iranians derived their ideas of sapta sindhavas and hapta hindu, respectively, from the Indus people themselves. For this, he utilized the concordance of the Finnish team and identified the occurrence of seven strokes in as many as 77 inscriptions. He also identifies 34 instances of seven strokes with a particular symbol, which, according to Kinnier Wilson, he identified as the Indus equivalent of the Sumerian symbol representing 'city' or 'enclosed space'.

In 1982, two books on Indus script decipherment were published by Krishna Rao and S.R. Rao. Krishna Rao begins with the proposition that Indus symbols were named with the first letter of a word representing that sound-word (phonogram). He identified the language of the inscriptions as Proto-Sanskrit and then proceeded to analyze the words, isolating their first phonetic syllable for naming characters. He used this formula to decipher the text of the Pashupati seal by referencing the names of the animals depicted therein. He starts from Buffalo and moves anti-clockwise. Therefore, he gets 'ma' from mahisha (buffalo) for the first letter, 'kha' from Khadga (rhinoceros) for the second, 'na' from nara for the third jar symbol, to the fish symbol he attributes 'sa' from shadri (elephant), followed by another 'na' for the jar sign. The reading he gets is Ma-Kha-Na-Sa-Na, which he takes to be a variant of makhahan—an epithet for Indra in Vedic literature.

S.R. Rao (1982) proposed that the Indus script has only 62 symbols, which were the basic signs, and the others were compounds formed by a combination of two or more basic symbols (ligatures) or variants of the same basic symbol (allographs). The nature of the script during the Mature Harappan period was pictographic, but it also functioned as a logo-syllabic system. He further reduced the signs to 20 based on the post-firing graffiti coming from the Late Harappan period, and thus the script became truly alphabetic. He credits the Harappan civilization with the invention of the alphabet.

After conducting an internal analysis of the symbols, Rao identified certain nuances in the script. He identifies the short strokes found in association with the linear signs as ‘vowel helpers’. The vertical lines found between symbols served as word separators. A single symbol could be used both for its determinative and phonetic value. He also isolated the functional aspect of each basic and compound symbol.

He assigned phonetic value of Semitic letters to the Indus alphabets and gave the reading of some of the texts as personal names, titles, and epithets. He also proposed that the language of the Indus script is closely related to Vedic Sanskrit.

1987 saw the entry of a scientist into the fold of Indus script decipherment. Kak reduced the number of symbols to just 39 basic characters, and the remainder he ascribes to ‘scribal variants, stylistic differentials and ligatures’. He identifies the Indus script as a forerunner to the later Brahmi script, and that since the latter has been deciphered, the former could also be read, proceeding from the known to the unknown. Kak carries out a statistical analysis of the frequency of the symbols both in the Indus corpus (Mahadevan’s) as well as in Brahmi. He also identifies morphological similarities between the two and attributes a tangible historical connection between them as the reason. He read the names of several Vedic gods in his decipherment.

Fairservis’s Jr. (1992) decipherment of the Indus script was deeply rooted in his conviction that the language of the script belonged to the Dravidian language family. Armed with this principle, he first identifies the basic iconography of the symbol. Next, he gleans the Dravidian vocabulary for a word that best suits the visual meaning conveyed by the symbol. In the final stage, he arrives at various meanings that can be assigned to the basic symbol and its subsequent variations.

Fairservis does not rely on computer-generated corpora that were available to him, but instead creates his own sets of grids for studying sign frequency, placement, and associations. Based on his grids, he identifies the Indus script as ideo-syllabic. He identifies 200 graphemes among the symbols, which swelled in number to 419 based on affixes and combinations. Even diacritical marks were used.

The value of Fairservis’s work was the hitherto first presentation of the Indus calendar. He opined that the ivory sticks found from Mohenjo-daro were, in fact, the visual representation of the calendars. These sticks had grooves, probably for sliding, in association with some of the Indus symbols. He identifies the calendar in use as a lunar one and even assigns names to them using his aforementioned decipherment methodology. He also believes that the number system of the Harappans was based on 8. This aligns well with the now-obsolete counting system of the early Dravidians.

In 1998, Wells listed upwards of 600 Indus symbols. He also downplays the claim that the Dravidian language family is a contender for the linguistic identity upon which the Indus script was based. Considering the nature of the script itself, he demonstrates that the Indus script is logo-syllabic. He compared his software-generated statistical findings on the Indus script with other known linguistic databases and concluded that the Indus writing system is linguistically based.

Institutional involvement in the search for the key to deciphering the Indus script in India started with the Tata Institute of Fundamental Research (TIFR) in 2007. A team of scientists began working with Iravatham Mahadevan, utilizing statistical and analytical tools to address issues related to the script. Mahadevan's concordance was the basis on which the tools were applied. These studies are supported by a financial grant from the Tata Trust. The team, comprising Rajesh P.N. Rao, Mayank Vahia, Nisha Yadav, Hrishikesh Joglekar, and others, has been working to decode the various facets of the Indus script and the underlying language through computational and statistical analyses. The probabilistic models created with the help of the proposed analyses will aid in identifying individual symbols as root words, suffixes, prefixes, other additives, or modifiers. This, in turn, may be helpful in identifying the grammatical bases of the underlying language and narrowing the field of the candidate linguistic family group from among the conflicting claims.

Steven Bonta, a linguist by training but with a deep-rooted interest in the Indus script puzzle, brought out his work on the script in 2010. He does not claim to have deciphered the script so much as assign meanings or values to some of the frequently occurring symbols and forward plausible hypotheses on the content of what he refers to as 'patterned inscriptions/texts.

In recent years, the field of decipherment has opened up to independent and amateur researchers who have occasionally contributed credibly to the growing corpus of literature on the subject.

Bahata Ansumali Mukhopadhyay is one such independent researcher with a professional background in computer applications. Through her numerous publications (Mukhopadhyay, 2018, 2019), she does not claim to decipher the script per se, but rather assigns meanings to symbols or their combinations that were intended to be conveyed through the inscribed texts. She identifies the inscribed seals, sealings, and tablets as data carriers, including taxation-related information such as tax rates, taxable commodities, Units of measurement in volumes/weights, etc., tax collectors, and modes of tax payment. She identifies the variants of bird-like logograms as representing various semi-precious stones, such as lapis lazuli, carnelian, and agate. Similarly, the fish-like logograms are identified as "fish-eye-

beads,” which were the most prized export to the ancient Near East. Moreover, she identifies the basic nature of the Indus script as logographic – symbols are content and functional morphemes, not phonograms for spelling.

The most recent work by Debasis Mitra’s team (2025) utilizes Artificial Intelligence (AI) to identify, extract, and store information from images of seals, specifically with respect to graphemes (from texts) and motifs. They have developed a custom-made application specifically for Indus script reading, which features an ‘Ancient Script Recognition Network’ (ASR-net) for recognizing graphemes and their sequences, and a ‘Motif Identification Network’ (MI-net) for identifying recurring iconographic motifs on seals. Their methodology is designed to provide a clear understanding of the script itself, its meaning, and its function.

Another independent researcher, Mahaveer H. Muhammad, claims to have deciphered the Indus script. He identifies the script as alphabetic in the service of a regional proto-Prakrit language form that he denotes as Sindhu Prakrit.

Bharat Rao ‘Yajna Devam’ – another independent researcher – claims to have deciphered the Indus script by treating it as a cryptogram (2024). In generic terms, a cryptogram is a message in a known language encoded in an unknown script. To crack the cryptogram, a substitution cipher is needed. In the present context, the inscriptions of the Indus script are the cipher text, and the underlying language of the script is the plain text. Yajna Devam utilizes cryptanalysis methods proposed by Shannon, combined with modern programming languages, for his decipherment.

He identifies the script form as ‘proto-abugida segmental’ where consonant-vowel sequences are written as a single unit, with the vowel sound indicated by a secondary mark or modification in the consonant’s glyph. The term ‘segmental’ has been used to denote that the graphemes (basic units) of the script represent individual sounds or segments of speech (consonants and vowels), rather than entire syllables.

Yajna Devam claims to have deciphered 500+ inscriptions of the Indus script. In doing so, he has come to the conclusion that the later-day Brahmi characters are standardized versions of Indus symbols and that there is a substantial legacy of Indus linguistic features and cultural elements in later periods as well.

Andreas Fuls started working in tandem with Bryan Wells for the development and maintenance of the Interactive Corpus of Indus Text (ICIT). Fuls was the computational brain behind the ICIT, looking after its development, programming, and maintenance. Having worked extensively on the Indus script with Wells, Fuls ventured out on his own with a two-volume publication (2022 and 2023) – *Corpus of Indus Inscriptions and Catalogue of Indus Symbols*.

Regarding the direction of reading, based on the data from his corpus, he states that 75% of 5659 texts are to be read from right to left. 4% has to be read from left to right (215 texts). Interestingly, 16 texts must be read from top to bottom, and 10 texts are written in boustrophedon. Also, for 14% of the texts, the direction of reading is unknown. Regarding the nature of the Indus script, Fuls reiterates that it is a mixed logo-graphic-syllabic system. He identifies 19-22% of the symbols as syllables and 78-81% as logograms.

Corpora and Concordances

Any attempt at deciphering a script depends on the database of the symbols. The more comprehensive and all-encompassing these databases are, the greater the chances of successful attempts at analytical research, which can subsequently lead to the formulation of hypotheses and their subsequent verification or refutation, partially or wholly.

Barton, in 1928, published what could reasonably be identified as the first attempt at compiling a sign list. His list had 124 symbols, including variants of basic symbols. He also found Sumerian parallels of 35 of these signs.

Langdon's list of symbols, published in 1931, listed the number as 288, with variations listed under the same number. Even at this early stage, his dataset comprised 566 examples of the Indus script, which contained approximately 2,500 symbols.

Hunter's publication (1932, 1934) is the earliest attempted concordance of the Indus script. He arranges the symbols, known to date, in tables and grids. This helps in identifying the basic symbols and those that are variants of the same motif. Hunter identifies the jar symbol as a terminal sign, and when it occurs within a text, the same is taken to denote a break between segments. The same conclusion was arrived at by the Soviet and Finnish teams through their computational analysis. He also makes a frequency count of the double depiction of the fish symbols, inclusive of variants, and concludes that they generally occur in the texts in the same sequential position in the same position. This led him to posit that the dual combination is a single word but may have different pronunciations depending on the dialect or euphony.

In 1939, Pran Nath in *Ancient Scripts of Harappan & Mohenjodaro* presents six charts of transcriptions of inscribed objects. His work is notable, not for the decipherment that he proposed, but for the listing of Indus symbols.

Dani, in his *Indian Palaeography* (1963), produces a sign list of the Indus script. He comes up with a large number of basic signs – 537. He also devises his own system

of categories for grouping these symbols.

The first truly comprehensive corpus and concordance of Indus symbols is by Mahadevan, published in 1977, titled “The Indus Script.” Texts, Concordance, and Tables. He identified 419 symbols based on the archaeological evidence available at the time (2911 artifacts; 3554 lines of text). With the help of a computer, he generated a font that matched the symbols to compile his concordance. He supplemented this database with information on the sites from which the individual evidences were recorded, the design (field symbol) on each seal, frequency, and distribution of each individual symbol, with reference to the texts, site, and field symbol. Mahadevan’s identification of the nature of the Indus script shifted with time from phonetic/ logographic/ word to ideographic. In 1980, the corpus was updated into a digital format under the nomenclature IDF80. Since 2021, the IDF80 corpus has been made available online through a web portal maintained by the Roja Muthiah Research Library, Chennai.

Another contemporaneous, slightly earlier series of concordances was produced by Parpola and his team throughout the 1970s and early 1980s (Koskeniemi, Parpola, and Parpola 1973; Koskeniemi and Parpola 1979, 1982). However, dissatisfied with the outcome of their decade-long work, Parpola invested in producing a complete photographic record of all known Indus texts. This ambitious UNESCO-backed project resulted in the publication of three volumes of the Corpus of Indus Seals and Inscriptions (Joshi & Parpola, 1987; Shah & Parpola, 1991; Parpola et al., 2010, 2020, 2022).

The latest attempt in this field is by Bryan Wells. His idea that came to fruition was the creation of an analytical computer program, the Interactive Concordance of Indus Texts (ICIT 5.2). The interactive software provides a compilation of statistics regarding the distribution of symbols and their use. It has been available online since October 2009 at www.epigraphica.de. This program is based on the Electronic Corpus of Indus Texts (ECIT). The corpus brings together examples of Indus texts, the artifacts on which these are found, and the archaeological context of the recovered artifacts. He also classified the textual evidence based on the nature of the artefact, text typology, and accompanying iconography, if any. The corpus also features a detailed sign list, including variants of the identified base symbols. Wells has published the ECIT as an Appendix (Wells, 2011). By 2015, he had increased the number of symbols to 694.

Fuls’ latest count of Indus symbols stands at 713 based on Wells’ ICIT database as of May 2023. This count is based on a total of 4,674 inscribed objects, which comprise 5,659 pieces of evidence of Indus texts and 19,869 recoded pieces of evidence of symbol occurrences. The database is available online through Ful’s website –

Epigraphica.de. The frequency of individual symbols is listed according to several parameters, including total frequency, frequency by text position, site, and artefact type.

All the corpora differ in their count of the repertoire of Indus symbols – Mahadevan 419, Parpola 386 plus 12 probable, Wells 694, Fuls 713. This is because of ambiguity in the identification of a motif as the base sign, or its variant – allograph or ligature. This is an aspect that is fraught with divergent subjective opinions and liable to constant revisions as new evidence and its plausible interpretations come to light.

With the help of the corpora, scholars have been able to find a relatively common ground in identifying basic symbols across the entire repertoire – approximately 150-200. Greater is the number of other symbols, most of which are compounds of the basic ones.

Some of the symbols are clearly pictorial, as expected. For example, the stick figure unambiguously represents a human; gender is inconclusive. Alternatively, it could be a divine figure as well. Another basic pictorial sign is the simple fish symbol – devoid of any additional features. As against the above, there are certain symbols where the pictorial motif is subject to biases – a circle or oval with intersecting lines could be a wheel or a chakra; a vertical with a number of horizontal lines could be seen as a comb, a harrow, or even a tree.

One of the biggest advantages of the above-mentioned corpora and concordances has been the generation of statistical data about the frequency of individual symbols, symbol combination while occurring in pairs or more, their relative positioning in a text, and their association, if any, with the graphical depictions, wherever present. Based on the above, the following conjectures have been made:

- The text segments in the majority of cases are one to three characters long and, in all probability, do not include grammatical affixes.
- Texts having a single motif probably represent a complete word in itself and can stand alone to convey the intended message/idea.
- The sequence of more than one character when repeated in the same combination in a longer text, the remainder of the text may be taken to constitute a self-contained segment of text.
- The relatively regular positioning of a particular symbol when occurring in a text gives some idea about its function. For example, it is generally seen that the ‘handled jar’ symbol occurs at the end of smaller texts – ‘segment terminal position’. Therefore, when occurring within a longer text, the text following the ‘handled jar’ can easily be identified as a separate section of the text.
- Some symbols have been, it appears, used interchangeably in certain cases. This

appears to be the case with the ‘arrow’ and the ‘handled jar’ motifs, which have been used to denote the ‘segment terminal position’.

Again, as with everything else in the ongoing saga of deciphering the Indus script, the above do not have universal acceptance. For example, Mahadevan (2000) interprets the ‘handled jar’ as person-number-gender mark based on the Dravidian nominal-singular-male/ nonmale gender distinction. Similarly, the fish symbols have been variously interpreted as a syllabic sign by Wells, an attributive-lexeme by Mahadevan, a unit of measurement by Bonta, ‘fish’, ‘star’, and ‘to shine or glitter’ by Parpola, based on the reading of ‘meen’ from the Dravidian language, and fish-eye-bead by Bahata Ansumali.



Working Group II

Theme: Survey, Documentation, Metadata Standards, and Digital Archiving

Coordinator

Prof. Ramesh C. Gaur
Dean (Admn.), IGNCA

Co-coordinator

Dr. Ajit Kumar
Director (CSL), Ministry of Culture, Govt. of India

Experts

Acharya Ajaysagarsuri
Senior Scholar, Koba Jain Gyan Mandir, Gujarat

Dr. Bhavanath Jha

Editor, Mahavir Mandir Publication Department, Patna

Dr. A. R. D. Prasad

Former Professor DRTC (ISI), Bangalore

Shri Sanjaya Singhal

Managing Trustee, Dharohar

Dr. P. M. Gupta

University Librarian, Central Sanskrit University, New Delhi

Dr. Kirtikant Sharma

Former Associate Professor, IGNCA

Dr. Parveen Babbar

Deputy Librarian, JNU

पाण्डुलिपि के सर्वेक्षण तथा- प्रस्तावना
पं. भवनाथ झा, संपादक, धर्मायण, महावीरमंदिर, पटना (बिहार)

यह हर्ष का विषय है कि केन्द्र सरकार के संस्कृति मन्त्रालय के द्वारा भारत की ज्ञान परम्परा के अन्तर्गत स्थित प्राचीन पाण्डुलिपियों के संरक्षण तथा उनके संपादन-प्रकाशन हेतु व्यापक प्रयास करने का निर्णय लिया गया है। इसके अन्तर्गत इसी उद्देश्य से 2003 में स्थापित राष्ट्रीय पाण्डुलिपि मिशन को विस्तारित करते हुए, उसे अधिक साधनसम्पन्न बनाते इस कार्य को आगे बढ़ाने का निर्णय लिया गया है। इसे आगे बढ़ाने के कार्य से जुड़े माननीय मन्त्री एवं अधिकारियों के प्रति आभार प्रकट करता हूँ।

दिनांक 11-13 सितम्बर 2025 को आयोजित “Reclaiming India’s Knowledge Legacy Through Manuscript Heritage” के लिए निर्मित निम्नलिखित विद्वानों के समूह में मुझे सर्वेक्षण एवं प्रलेखन विषय पर प्रस्ताव देने का भार दिया गया है-

मैं 1998 ई. से व्यक्तिगत रूप से पाण्डुलिपि के सर्वेक्षण तथा अप्रकाशित पाण्डुलिपियों के सम्पादन तथा प्रकाशन से जुड़ा रहा हूँ। कामेश्वर सिंह दरभंगा संस्कृत विश्वविद्यालय के लिए मैंने 18 पाण्डुलिपियों का मैथिली लिपि से लिप्यन्तरण भी किया है। पटना संग्रहालय में स्थापित राष्ट्रीय पाण्डुलिपि मिशन के संसाधन केन्द्र के सर्वेक्षक समूह से मेरा वैयक्तिक सम्पर्क रहा है। उनके क्रियाकलापों और उनकी कठिनाइयों से मैं अवगत होता रहा हूँ, अतः इतने वर्षों के अनुभव के आधार पर सर्वेक्षण तथा प्रलेखन विषय पर अपना मन्तव्य निम्नवत् रख रहा हूँ।

सर्वेक्षण की आवश्यकता

पाण्डुलिपियाँ भारतीय ज्ञान-परम्परा के लिखित स्रोत हैं, जो भारत की विभिन्न संस्थाओं के साथ-साथ व्यक्तिगत अधिकार में घरों में उपलब्ध हैं। ऐसी स्थिति में इनकी कुल संख्या का आकलन करना आज भी कठिन है। यद्यपि राष्ट्रीय पाण्डुलिपि मिशन के द्वारा लगभग 20 वर्षों में अनेक सर्वेक्षण कराए गये हैं, किन्तु अभी गहन सर्वेक्षण की आवश्यकता है, ताकि हम सभी पाण्डुलिपियों की वास्तविक स्थिति का पता लगा सकें।

इस कार्य हेतु यहाँ पर तीन शीर्षकों के अन्तर्गत विचार प्रस्तुत है-

1. पाण्डुलिपियों के प्राप्ति-स्थान
2. सर्वेक्षण की व्यावहारिक एवं सैद्धान्तिक समस्याएँ
3. समाधान

पाण्डुलिपियों के स्थान

भारत के विभिन्न भू-भाग में पाण्डुलिपियों के प्राप्ति-स्थान दो प्रकार के हैं-

1. संस्थागत संग्रह तथा 2. व्यक्तिगत संग्रह ।

संस्थागत संग्रह

भारत में संस्थागत संग्रह में सर्वाधिक पाण्डुलिपियाँ होने का अनुमान है। प्राचीन काल की बड़ी-बड़ी रियासतों के पास अपना-अपना पुस्तकालय था, जिसमें राजा या जमींदार की रुचि के अनुसार पाण्डुलिपियों का संग्रह किया गया था। संस्थागत संग्रह की विशेषता है कि अनेक विषयों तथा अनेक लिपियों की पाण्डुलिपियाँ हमें एकत्र मिल जाती हैं। इस संस्थागत संग्रह के दो प्रकार हैं-

1. सार्वजनिक संस्था तथा 2. निजी संस्था

सार्वजनिक संस्था

सार्वजनिक संस्था से यहाँ तात्पर्य है कि जहाँ संस्थागत प्रकृति का संग्रह है और उस पर नियन्त्रण किसी एक व्यक्ति का न होकर सरकार या किसी समिति का है। सरकारी सहायता प्राप्त संस्थाओं की पाण्डुलिपियों तक पहुँचने के लिए हमें भिन्न पद्धति अपनानी होगी, किन्तु समिति द्वारा संचालित संग्रह तक पहुँचने के लिए बिल्कुल अलग पद्धति होगी, अतः हम सार्वजनिक संस्था को भी पाण्डुलिपि सर्वेक्षण की दृष्टि से दो भागों में बाँट सकते हैं-

1. सरकारी तथा 2. गैर-सरकारी

सरकारी संस्थाएँ

सरकारी संस्थाएँ चूँकि एक प्रशासन तंत्र के अन्तर्गत स्थापित हैं अतः इनके अधिकार में संकलित पाण्डुलिपियों के सर्वेक्षण के लिए इस तंत्र की सहायता अपेक्षित होती है। इसके अन्तर्गत शिक्षण संस्थान यथा- विश्वविद्यालय, शोध संस्थान आदि हैं। भारत में स्थापित विभिन्न विश्वविद्यालयों से सम्बद्ध अनेक स्थानीय महाविद्यालय हैं। इनके पुस्तकालयों में स्थानीय स्तर पर स्वेच्छा से अतीत में बहुत सारी पाण्डुलिपियाँ संकलित की गयी हैं। विशेष रूप से संस्कृत विश्वविद्यालयों से संबद्ध महाविद्यालयों तथा विद्यालयों को हमें ध्यान में रखना चाहिए। बिहार राज्य की स्थिति को देखते हुए चर्चा करें तो व्यक्तिगत संग्रह से भी अधिक दयनीय स्थिति इनमें संकलित पाण्डुलिपियों की है। इसी प्रकार, सार्वजनिक पुस्तकालयों की भी अच्छी स्थिति नहीं है। सरकारी नियंत्रण होने के बाद भी ये बंद पड़े होते हैं तथा यहाँ से पुस्तकों के साथ पाण्डुलिपियाँ तक या तो व्यक्तिगत अधिकार में आ चुकी हैं या अंतरराष्ट्रीय बाजार में बेच दी गयी हैं। इन पाण्डुलिपियों का तत्काल सर्वेक्षण कर संरक्षण करने की आवश्यकता है। राज्य सरकार के अन्तर्गत पुरातात्विक संग्रहालय भी पाण्डुलिपि की दृष्टि से

महत्त्वपूर्ण हैं। बिहार प्रदेश की बात करें तो चन्द्रधारी मिथिला संग्रहालय, दरभंगा, पटना संग्रहालय, पटना, में हजारों पाण्डुलिपियाँ एकत्र हैं। इनके सर्वेक्षण तथा संरक्षण के लिए हमें तत्काल सरकारी तंत्र का उपयोग कर कार्य करना चाहिए।

गैरसरकारी (किसी कमीटी या ट्रस्ट द्वारा संचालित)

मठ, मन्दिर, गुरुद्वारा, खानकाह, आदि धार्मिक स्थल, निजी पुस्तकालय- ये सब गैर-सरकारी संस्थाएँ हैं। इनमें पाण्डुलिपियों का विशाल भण्डार है। बोधगया के शांकर मठ में एक ऐसा संग्रह मिला है, जिसके बारे में वहाँ के लोगों को भी सूचना नहीं थी। कहा जाता है कि लगभग 200 वर्षों से वह कमरा बंद था, एक गुप्त द्वार था, जो हाल में उद्घाटित हुआ है। वहाँ लगभग 7000 पाण्डुलिपियाँ एकत्र हैं। ऐसे संग्रह में हम एक विशेष दार्शनिक विचार धारा के ग्रन्थ पाते हैं। निजी पुस्तकालयों की स्थापना भी किसी समिति के अंतर्गत हुई है, जिन्हें सरकारी सहायता प्राप्त नहीं है, किन्तु किसी समिति के द्वारा संचालित हैं। इनका सर्वेक्षण होना चाहिए तथा उचित माध्यम का उपयोग कर यहाँ की पाण्डुलिपियों का संरक्षण अपेक्षित है।

निजी संस्थाएँ

निजी संस्था के अन्तर्गत हम ऐसे पुस्तकालय को लेंगे, जो किसी व्यक्तिविशेष के अधिकार में है। वहाँ न तो कोई समिति है, न ही किसी प्रकार से सरकारी नियंत्रण है। भारत में जमींदारी उन्मूलन के बाद भी बड़ी-बड़ी रियासतों में निजी पुस्तकालयों का अस्तित्व हम पाते हैं। ये उस राज-परिवार के नियंत्रण में है। आज वहाँ न तो कोई कर्मचारी है, न ही वह कभी खुलता है। राज-परिवार के सदस्य विदेशों में या देश के महानगरों में रहते हैं। यदि हम उन रियासतों की सूची के अनुसार उनसे सम्पर्क कर अनुमति लेते हैं, तो वहाँ की पाण्डुलिपियाँ ज्ञान-परम्परा तथा कला की दृष्टि से महत्त्वपूर्ण होगी।

व्यक्तिगत अधिकार में

उपर्युक्त निजी संस्था में संकलित पाण्डुलिपियों के अतिरिक्त व्यक्तिगत घरों में पूर्वजों के द्वारा संकलित पाण्डुलिपियों का विशाल भण्डार है, किन्तु इनकी प्रकृति भिन्न है। यहाँ हमें विषयानुसार पाण्डुलिपियाँ मिलती हैं। जिस परिवार के पूर्वज जिस विषय के विद्वान् थे, उनके संकलन में एक एक विषय की पाण्डुलिपियाँ हमें मिलती हैं। अतः इन्हें हम पारिवारिक पाण्डुलिपियाँ कह सकते हैं। आज की स्थिति में हमें स्थानीय स्तर पर यह पता लगाना होगा कि किस व्यक्ति के पूर्वज शास्त्रों के विद्वान् थे। उनकी वंशावली के आधार पर उचित गृहस्वामी तक पहुँच सकते हैं, जहाँ पाण्डुलिपियों की संभावना हो सकती है।

पुराने जमींदार परिवार में हमें अधिक बड़े संग्रह की आशा है। अतीत में अनेक छोटे-छोटे जमींदार भी

शिक्षा-प्रेमी होते थे और विभिन्न शास्त्रों के पाण्डुलिपियों का संग्रह किया करते थे। पारिवारिक संग्रह की अपेक्षा यहाँ अधिक महत्त्वपूर्ण पाण्डुलिपियाँ मिलने की संभावना है।

इसके साथ पंजी-कारों/तीर्थपुरोहितों के घरों में प्राप्त वंशावली, तीर्थयात्रियों के विवरण आदि अभिलेख तत्कालीन इत्हास के महत्त्वपूर्ण स्रोत हैं। मिथिला की बात करें तो वहाँ का पंजी-प्रबन्ध व्यवस्थित है, जिसमें लगभग 14वीं शताब्दी से वंशावली संकलित है। इनकी पाण्डुलिपियाँ भी इतिहास के महत्त्वपूर्ण स्रोत हैं। मधुबनी जिला के सौराठ, सहरसा जिला का महिषी आदि गाँवों में पंजीकार हैं, जिनके घरों में इस प्रकार के संग्रह पाये जाते हैं। इनके अतिरिक्त स्थानीय स्तर पर हमें अन्य घरों से भी सूचनाएँ मिल सकेगी, जिनका सर्वेक्षण अपेक्षित होगा।

सर्वेक्षण की व्यावहारिक एवं सैद्धान्तिक समस्याएँ

ऊपर हमने पाण्डुलिपि के प्राप्ति-स्थानों का परिचय दिया है। सर्वेक्षण के सन्दर्भ में सभी स्थानों की अपनी-अपनी और अलग-अलग समस्याएँ हैं। राष्ट्रीय पाण्डुलिपि मिशन ने भी सर्वेक्षण के लिए बहुत उपाय किए थे। बहुत जगहों पर उन्हें काम करने में सफलता भी मिली, किन्तु अनेक संग्रहालयों से भी उन्हें खाली हाथ लौट जाना पड़ा। घर के द्वार तक पहुँचे पर उन्हें गृहस्वामी के द्वारा बताया गया कि हमारे घर में कोई पाण्डुलिपि नहीं है। ऐसा स्थिति क्यों उत्पन्न हुई, इस पर विचार आज अपेक्षित है और यह अपेक्षा की जाती है कि ज्ञान भारतम् मिशन के द्वारा उनका समाधान किया जाए।

यूरोपीयन शासन काल से ही पाण्डुलिपि को लेकर कोई अधिनियम नहीं बन सका। जिस प्रकार, भारतीय पुरातत्त्व सर्वेक्षण के लिए अलग संस्था बनी, उसका अधिनियम बना, जो हर जगह लागू होता था, उसका उल्लंघन करने वालों के लिए सजा का प्रावधान किया गया, ऐसा कोई अधिनियम पाण्डुलिपि के विषय में नहीं बन सका। राष्ट्रीय पाण्डुलिपि मिशन की भी जो स्थापना हुई तो अधिनियम का अभाव रहा। इसे हमेशा एक मिशन के रूप में रखा गया। पाण्डुलिपि को राष्ट्रीय धरोहर घोषित करने की बात उठती रही, लेकिन आगे काम नहीं हुआ।

फलतः राष्ट्रीय पाण्डुलिपि मिशन के समानान्तर अनेक संस्थाएँ गठित हुईं और मनमाने ढंग से काम करने लगी। ब्रिटिश लाइब्रेरी के Endangered Archives Programme के तहत बंगाल, आसाम तथा नेपाल की पाण्डुलिपियों का डिटाइजेशन कराया गया। ई-गंगोत्री ने भी पुस्तकों तथा पाण्डुलिपियों का डिजिटाइजेशन कराया। एक संस्था Indian National Trust for Art and Cultural Heritage ने भी पाण्डुलिपियों के संरक्षण पर कार्य किया है।

अब समस्या है कि किसी विश्वविद्यालय के कुलपति अथवा गाँव का एक गृहस्वामी किसे अपनी पाण्डुलिपि दिखाये? उन्हें कैसे विश्वास होगा कि इस संस्था के सर्वेक्षक दल के सदस्य पाण्डुलिपियों के साथ काम करने के लिए कुशल हैं? यदि किसी संस्था के अधिकारी अपनी पाण्डुलिपियों को नष्ट-भ्रष्ट करने से बचाने की मानसिकता रखते हुए सर्वेक्षण की अनुमति नहीं देते हैं, तो इसमें हम उनकी गलती नहीं कहेंगे।

पाण्डुलिपि तस्करी भी सर्वेक्षण की एक समस्या है। लोगों के मन में यह बैठ गया है कि हमारी पाण्डुलिपियों का मूल्य अंतरराष्ट्रीय बाजार में लाखों-करोड़ों रुपयों में है। व्यक्तिगत संग्राहक इसे गोपनीय समझने लगे हैं। हालांकि गोपनीयता का एक अन्य पारिवारिक कारण भी है। पारिवारिक संपत्ति विभाजन में यदि पाण्डुलिपियाँ किसी एक हिस्सेदार के घर में रह गयीं तो वे उसे अन्य हिस्सेदार से छुपाने का प्रयास करते हैं। ये सब धरालत की बातें हैं, जो सर्वेक्षण के बाधक रहे हैं।

आजकल डिजिटल पाण्डुलिपि की भी तस्करी होने लगी है। सर्वेक्षक दल तस्कर गिरोह के हैं या वास्तविक सर्वेक्षक हैं, इसका विश्वास धारक को कैसे मिलेगा? पटना संग्रहालय के परिसर में राष्ट्रीय पाण्डुलिपि मिशन का संसाधन केन्द्र खुला तो उसके सर्वेक्षक के पास राष्ट्रीय पाण्डुलिपि मिशन का परिचय पत्र तक नहीं होता था। पटना संग्रहालय ने अपने परिसर में निःशुल्क प्रवेश के लिए जो परिचय पत्र बना कर दिया था, उसी को दिखा कर यहाँ के सर्वेक्षक कार्य करते थे। उस केन्द्र के द्वारा सर्वेक्षित लगभग 68,000 पाण्डुलिपियों का डाटाबेस पाण्डुलिपि पटल पर उपलब्ध है। ऐसे राष्ट्रीय पाण्डुलिपि मिशन के संसाधन केन्द्र पटना को संस्कृति मंत्रालय द्वारा परिचय पत्र जारी ना करने का क्या कारण था, यह समझ से परे हैं। अतीत में राष्ट्रीय पाण्डुलिपि मिशन की जो भी बाध्यता रही हो, संवैधानिक सीमाएँ रही हो, जिसने सर्वेक्षण के कार्यों को प्रभावित किया हो, उनका समाधान आज वैधानिक स्तर पर अपेक्षित है।

समाधान:

राष्ट्रीय पाण्डुलिपि बिल अधिनियम (NMA)

जैसा कि समाचार-पत्रों के माध्यम से सूचना मिली है कि राष्ट्रीय पाण्डुलिपि अधिनियम के लिए सन् 2023 से ही तैयारी चल रही है और इसका बिल तैयार हो चुका है, उसे शीघ्र कार्यान्वित करना की आवश्यकता है, जो हमें सर्वेक्षण में हुई कठिनाइयों का सैद्धान्तिक समाधान देगा।

इसमें पाण्डुलिपि को ज्ञान-पाठ्य के रूप में परिभाषित करने की आवश्यकता है। हमें यह मानना चाहिए कि पुरातात्विक सामग्री से कहीं अधिक महत्त्व उसके पाठ का है। जो ग्रन्थ पूर्व संपादित भी हो चुके हैं, उनकी भी आज यदि अनेक पाण्डुलिपियाँ मिलती हैं, तो पाठों का पुनरीक्षण आवश्यक है। अतः पाण्डुलिपियाँ अन्ततः एक knowledge text है। ज्ञान भारतम् को इसके पाठ पर समुचित ध्यान देना चाहिए। वर्तमान में पाण्डुलिपि की वैधानिकता मात्र एक पुरातात्विक सामग्री के रूप में है। भारतीय पुरातत्त्व सर्वेक्षण के The Ancient Monuments And Archaeological Sites And Remains Act, 1958 में अनुच्छेद 2, (b), (i) में पाण्डुलिपि को भी पुरातात्विक सामग्री के रूप में सम्मिलित किया गया है और जहाँ पर पुरातात्विक सामग्रियों के पंजीकरण की बात की है, तो वहाँ पाण्डुलिपि को भी अन्तर्भूत मान लिया जा सकता है।

प्रस्ताव 1.

इस प्रकार का एक नियम 'ज्ञान भारतम्' के लिए अधिनियम के तहत बनाया जा सकता है, ताकि किसी भी संस्था में या व्यक्तिगत घरों में यदि कोई पाण्डुलिपि है, तो उनके पास रह तो सकती है, किन्तु उसका पंजीकरण कराना अनिवार्य हो जाएगा। इस पंजीकरण का अधिकार ज्ञान भारत के अतिरिक्त किसी भी संस्था को नहीं रहेगी। जो सामग्रियाँ पंजीकृत रहेगी, उसका प्रमाणपत्र धारक को दिया जाएगा। पंजीकरण के इस अधिकार को जारी रखने के लिए ज्ञान भारत मिशन को एक 'आयोग', 'प्राधिकरण' या 'स्वतंत्र संस्था' घोषित करना होगा। साथ ही, इसके समानान्तर कार्य करने वाली संस्था पर प्रतिबन्ध लगाना होगा। उन्हें यदि काम करना है, तो पहले ज्ञान भारतम् से उसका संबंधन करवायें और किए गये कार्य का प्रतिवेदन दें। इस प्रकार हमें सरकारी नियंत्रण में संकलित पाण्डुलिपि संधारण केन्द्रों तथा व्यक्तिगत संग्रह में संकलित पाण्डुलिपि स्वामियों को विश्वास दिलाना होगा कि भारतीय नागरिक होने का यह कर्तव्य है। इससे संस्थागत संग्रह की पाण्डुलिपियों के सर्वेक्षण का समाधान हो जाएगा। व्यक्तिगत संग्रह में स्थित पाण्डुलिपियों के लिए भी हमें के लिए हमें उपर्युक्त समाधान पचास प्रतिशत काम करेगा। प्रबुद्ध जन इसके आधार पर अपनी पाण्डुलिपि का पंजीकरण करा लेंगे।

प्रस्ताव 2 - 'पाण्डुलिपि संरक्षण पुरस्कार'

ज्ञान भारतम् 'पाण्डुलिपि संरक्षण पुरस्कार' की घोषणा करे। जो व्यक्ति इतने दिनों से धरोहर के रूप में अपने घरों में पाण्डुलिपि का संरक्षण करने का कार्य निभा रहे हैं, उन्हें पुरस्कृत किया जाना चाहिए।

1. यह पुरस्कार केवल व्यक्तिगत संग्रह की पाण्डुलिपियों के लिए दिया जाएगा।
2. पाण्डुलिपि की संख्या तथा उसकी गुणवत्ता के आधार पर संधारक को पुरस्कृत किया जाएगा।
3. संधारक एक निर्धारित अवधि में अपने घर में संकलित पाण्डुलिपियों की सूची Online उपलब्ध करावेंगे।
4. यदि उन्हें विवरणी स्वयं बनाने में कोई समस्या होती है तो अपना पूरा पता देकर ज्ञान भारत मिशन से सहायता की माँग करेंगे। स्थानीय संसाधन केन्द्र के कुशल सर्वेक्षक उनके घर जाकर उनकी मदद करेंगे।
5. इस पुरस्कार योजना को मंत्रालय की मान्यता प्राप्त होगी।

इस योजना से सर्वेक्षण में आने वाली बहुत बाधाएँ बहुत हद तक समाप्त हो जाएगी। आज भी पुरस्कार की राशि का लोभ उन्हें पाण्डुलिपि के महत्त्व को समझने के लिए बाध्य करेगा। साथ ही, 'संस्कृति मंत्रालय' के अधीन 'ज्ञान भारतम्' इस प्रकार का कार्य कर रहा है, इसका प्रचार-प्रसार

धरातल पर होगा। इससे 'ज्ञान भारतम्' की वैधानिकता तथा पाण्डुलिपि के लिए एकमात्र संस्था होने की बात दूर-दराज के क्षेत्रों तक फैलेगी। लोगों में 'ज्ञान भारतम्' का परिचय पत्र लेकर आने वाले सर्वेक्षक दल के प्रति विश्वास की भावना उत्पन्न होगी, जिससे घर में पाण्डुलिपि रहते हुए भी नकार देने की असहज स्थिति नहीं आने की संभावना है।

प्रस्ताव 3

सर्वेक्षण के लिए ज्ञान भारतम् को सरकारी तंत्र का सहारा लेना चाहिए। आज सरकारी संस्थागत संग्रह की बात करें तो सर्वाधिक पाण्डुलिपियाँ शिक्षा मन्त्रालय के अधीन है। विश्वविद्यालय, केन्द्रीय संस्कृत विश्वविद्यालय, महाविद्यालय, विद्यालय, शोध-संस्थान, पुस्तकालय, राष्ट्रीय एवं प्रान्तीय अभिलेखागार, राष्ट्रीय एवं प्रान्तीय संग्रहालय- ये सब बड़े-बड़े संग्राहक केन्द्र हैं। विभागीय रूप में इनके साथ पाण्डुलिपियों को विशेष दर्जा देने पर सहमति लेनी चाहिए। इन सरकारी पाण्डुलिपि संग्रहों के लिए सर्वेक्षण दल को विभागीय सहमति पत्र उपलब्ध कराना चाहिए। शिक्षा मंत्रालय विश्वविद्यालयों को आदेश करेगा, विश्वविद्यालय अपने महाविद्यालयों को आदेश करेगा। इस प्रकार सर्वेक्षण के लिए केन्द्रीय मन्त्रालय एवं राज्य मन्त्रालयों के स्तर पर एक अनुबन्ध होना चाहिए।

गैर-सरकारी संस्थाओं एवं व्यक्तिगत संग्रह के लिए 'ज्ञान भारतम्' को अपनी अर्हता, वैधानिकता तथा पाण्डुलिपि के लिए एकमात्र संस्था होने की घोषणा करनी चाहिए। साथ ही, लोगों में पाण्डुलिपि के प्रति जागरूकता लाना आवश्यक है।

पाण्डुलिपि संसाधन केन्द्रों को मजबूत बनाना

राष्ट्रीय पाण्डुलिपि मिशन के द्वारा पूरे देश में संसाधन केन्द्र (MRC) तथा संरक्षण (MCC) केन्द्र खोले गये हैं। इनमें से बहुत सारे केन्द्र आज बंद पड़े हैं। वास्तव में पाण्डुलिपियों का सर्वेक्षण, प्रलेखन तथा संरक्षण एक दूसरे से इतने निकट हैं कि केन्द्रों के इन दो प्रकारों की आवश्यकता ही नहीं है। वास्तव में देखा जाए तो पाण्डुलिपि पर कार्य की यह क्रमिक दिशा है-

सर्वेक्षण (Survey) > संरक्षण (conservation) > प्रलेखन (Documentation) > डिजिटाइजेशन Digitization) > आर्काइविंग (Archiving)>पाठ संपादन (Editing) >प्रकाशन Publication.

इसके क्रम में परिवर्तन करने से हम लक्ष्य को नहीं पा सकेंगे। कई बार स्थिति आती है कि बिना संरक्षण किए हम पाण्डुलिपि को खोल तक नहीं पाते हैं। 19वीं शती की कम्पनी निर्मित स्याही से लिखी पाण्डुलिपियों में जंग लगने के कारण पन्ने एक-दूसरे से चिपक जाते हैं तथा उन्हें अलग करने पर पत्र जालीदार हो जाते हैं। इनका प्रलेखन बिना संरक्षण का संभव नहीं है। यह संरक्षण विज्ञान का विषय है तो प्रलेखन के लिए लिपि, भाषा एवं शास्त्र आवश्यकता होती है। अतः प्रलेखन के लिए जो दल हम

तैयार करें, उसके साथ संरक्षण के विशेषज्ञ भी होना चाहिए। साथ ही, उनके साथ भाषा, क्षेत्रीय लिपि तथा शास्त्र के भी जानकार होना चाहिए।

प्रत्येक पाण्डुलिपि केन्द्र पर एक समन्वयक का अनुबन्ध के आधार पर पद दिया जा सकता है, जो क्षेत्रीय लिपि, भाषा तथा पाण्डुलिपि के विशेषज्ञ हों, उनमें संपादन तक करने का अनुभव होना चाहिए। साथ ही, उसी केन्द्र पर संरक्षण की भी व्यवस्था हो। वे अपने निर्देशन में प्रलेखन के दल से कार्य करावेंगे तथा किए गये कार्यों का निरीक्षण करेंगे। उस केन्द्र पर डिजिटाइजेशन की भी सुविधा होनी चाहिए तथा आर्काइविंग के लिए भी वहाँ एक्सेस मिलना चाहिए।

इस प्रकार, पाण्डुलिपि संसाधन केन्द्र (MRC) तथा पाण्डुलिपि संरक्षण केन्द्र (MCC) दोनों को एक साथ कर देना अपेक्षित होगा और एक निदेशक/समन्वयक को इसका पूर्ण उत्तदायी बनाकर स्थापित करना अपेक्षित होगा। वे स्थानीय विद्वानों के साथ बैठक कर सर्वेक्षित तथा प्रलेखित पाण्डुलिपियों में से अप्रकाशित ग्रन्थों का चुनाव कर उनके प्रकाशन की दिशा में भी चिन्तन करेंगे तथा यथासम्भव सम्पादित कराकर उसकी डिजिटल कापी केन्द्रीय कार्यालय को भेज देंगे।

इस प्रकार, 'ज्ञान भारतम्' के क्षेत्रीय ईकाइयों के गठन से कार्य में अपेक्षित सफलता मिलेगी तथा अभी तक जो सर्वेक्षण प्रलेखन आदि हो चुके हैं, उसका भी पुनरीक्षण होगा।

नकारात्मक अपवाह के विरुद्ध जागरूकता

युवा पीढ़ी में पाण्डुलिपि के महत्त्व को लेकर जागरूकता आवश्यक है। बहुत कुछ नकारात्मक बातें समाज में फैलायी गयी हैं। इन पंक्तियों का लेखक एक विश्वविद्यालय के भूतपूर्व कुलपति तथा भारतीय प्रशासनिक सेवा के उच्च अधिकारी के मुँह से यह सुनकर चौंक गया कि अब जो पाण्डुलिपियाँ हैं वे नकली हैं, उनमें परवर्ती काल में बहुत जोड़े गये हैं। दूसरी नकारात्मकता है कि लेखक के हाथ की लिखी पाण्डुलिपि ही महत्त्वपूर्ण हैं। मैं बिहार का हूँ, वहाँ एक बात लोग बरबस कहते सुने जाते हैं- 'हमारा जो कुछ था, नालंदा के साथ जल गया। अब बचाने के लिए कुछ बचा ही नहीं।' लगभग 150 वर्षों से कुछ धार्मिक संगठन इस प्रकार की भ्रान्तियाँ फैलाने में लगे हैं। इनके विरुद्ध जागरूकता लाना भी आवश्यक है, ताकि हमारी युवा पीढ़ी पाण्डुलिपियों के वास्तविक शैक्षिक महत्त्व को समझ सके। साथ ही, यह सकारात्मक भाव मन में ला सकें कि अभी भी बहुत कुछ है, जिसके प्रति हमारा कर्तव्य बनता है।

युवा पीढ़ी को पाण्डुलिपि अध्ययन के प्रति रुचि जगाना

युवा पीढ़ी को जहाँ रोजगार मिलेगा, उसकी ओर वे प्रेरित होंगे। पाण्डुलिपि अध्ययन से रोजगार की संभावना हमें तलाशनी होगी। इन्हें हम निम्न प्रकार से रोजगार दे सकते हैं-

1. लिपिवाचक (Decipherment Expert) के रूप में- इसके लिए ज्ञान भारतम् के वेबसाइट पर एक लिपिवाचकों की राष्ट्रीय पंजी बनानी चाहिए। जो लोग पाण्डुलिपियों को पढ़कर उन्हें

आधुनिक लिपियों में यूनिकोड में कम्प्यूटर टाइपिंग करने के प्रति रुचि रखते हों, वे अपना परिचय देकर स्रोत-लिपि तथा लक्ष्य-लिपि का विवरण देकर अपना नाम पंजीकृत करा सकते हैं। वे अपना भाषा-ज्ञान तथा शास्त्र-ज्ञान का भी विवरण देंगे। इस राष्ट्रीय पंजी के आधार पर उन्हें अनुबन्ध के माध्यम से कार्य मिलते रहेंगे। इस लिप्यन्तरण तथा कम्प्यूटर टाइपिंग से उन्हें आर्थिक लाभ होगा। भाषा-साहित्य का अध्ययन करने वाले युवा-युवतियों के लिए यह एक सुनहरा अवसर होगा, जिसमें घर पर रह कर या खाली समय में काम कर वे अर्थोपार्जन कर सकते हैं। इस प्रकार का कार्य करते-करते वे भविष्य के लिए अच्छे संपादक भी बन जाएंगे। ऐसे लोगों को यदि किसी प्रकार के निर्देशन की अपेक्षा होगी तो कार्यशालाओं के माध्यम से उनकी सहायता की जा सकती है। भारत या भारत से बाहर के किसी शोधार्थी या संपादक को विशेष लिपि की पाण्डुलिपि का वाचन-लिप्यन्तरण कराना होगा, तो वे संगत लिपि तथा भाषा के विशेषज्ञ लिपि-वाचकों से सम्पर्क कर उन्हें पाण्डुलिपि की डिजिटल प्रति भेज कर, वार्ता के माध्यम से पारिश्रमिक तय कर, कार्य करा सकेंगे। इससे घर पर रहने वाली महिलाओं को स्वरोजगार का भी अवसर मिलेगा, भाषा-साहित्य के अध्येता बेरोजगार युवकों को घर पर रहते हुए सम्मानजनक कार्य का अवसर तथा रोजगार मिलेगा। इस एक Online Application से बहुमुखी लाभ होंगे। सके निर्माण के लिए भारतीय भाषा संस्थान, मैसूर के अंतर्गत स्थापित राष्ट्रीय अनुवाद मिशन की पंजी देखी जा सकती है। यहाँ अभी तक इस पर 8996 अनुवादक पंजीकृत हो चुके हैं। <https://www.ntm.org.in/languages/english/login.aspx> संरक्षण-विशेषज्ञ के रूप में- पाण्डुलिपि के संरक्षण में कुशल व्यक्तियों की हमेशा आवश्यकता होती रहेगी। युवाओं के लिए यह रोजगार का बड़ा अवसर होगा। उन्हें विशेष प्रशिक्षण देकर अनुबन्ध के आधार पर जोड़ा जा सकता है।

2. सर्वेक्षण के लिए- स्थानीय व्यक्तियों को, स्थानीय पत्रकारों को इस सर्वेक्षण में अनुबन्धित किया जा सकता है, जो पाण्डुलिपि केन्द्रों को व्यक्तिगत संग्रह में स्थित पाण्डुलिपियों की सूचना देंगे तथा धारकों को अपने अधिकार में स्थित पाण्डुलिपियों के संरक्षण तथा प्रलेखन/पंजीकरण के लिए प्रोत्साहित करेंगे। इसके लिए उन्हें किए गये कार्यों के आधार पर राशि का भुगतान किया जाएगा। ये एक प्रकार से एजेंट होंगे। आज संचार माध्यम सुलभ होने के कारण सुदूर ग्रामीण क्षेत्र में भी स्थानीय पत्रकार जन-संपर्क साधे हुए रहते हैं। यदि ज्ञान भारतम् के द्वारा पाण्डुलिपियों की खोज के लिए बी अनुबन्ध किया जाता है तो सर्वेक्षण के लिए प्रमुख घटक सिद्ध होंगे।
3. प्रलेखन के लिए अनुबन्ध देकर- लिपिवाचकों की राष्ट्रीय पंजी में से स्थानीय लिपिवाचकों का चुनाव कर अनुबन्ध के आधार पर ज्ञान भारतम् के स्थानीय केन्द्रों के द्वारा चुनाव किया जाएगा, उनकी टीम प्रलेखन के लिए घर-घर जाकर कार्य करेगी तथा स्थानीय केन्द्र को अपना रिपोर्ट सौंपेगी।
4. इस प्रकार पाण्डुलिपि-अध्ययन को रोजगार से जोड़ने पर युवा पीढ़ी में इसके अध्ययन के प्रति

रुचि जगेगी, जिससे ज्ञान भारतम् के अनेक कार्य सिद्ध होंगे।

पाण्डुलिपि के लिए AI का प्रयोग

एकल गये डाटा के विश्लेषण के लिए AI का प्रयोग करने से कार्य में प्रगति होगी, किन्तु आज AI इस स्थिति में नहीं है कि हम उस पर निर्भर हो जाएँ। हमें भाषा, लिपि तथा शास्त्र के विशेषज्ञ को अनुबंधित करना ही होगा।

विशेष रूप से व्यक्तिगत संग्रह की पाण्डुलिपियाँ अस्त-व्यस्त रहती हैं। एक पाण्डुलिपि के पृष्ठ दूसरी पाण्डुलिपि में घुसे होते हैं। एक ही लिपिकार की लिखी, एक ही आकार के आधार (कागज, तालपत्र, तूलपट्ट आदि) पर लिखी पाण्डुलिपि एक-दूसरी से इस प्रकार मिश्रित हो जाती है कि प्रलेखन से पहले इन्हें सुलझाना सबसे बड़ा काम होता है। अक्सर व्यक्तिगत संग्रह की पाण्डुलिपियों में हमें एक लिपिकार की अनेक पाण्डुलिपियाँ एकल मिलती हैं। उन्हें जब तक व्यवस्थित न कर लिया जाए तब तक आगे कार्य सम्भव नहीं होगा। इस व्यवस्थापन के लिए भी, डिजिटाइजेशन के बाद AI के उपयोग का सुझाव आया था, जिसे स्वीकार करना कठिन होगा। विभिन्न लिपिकार एवं आकार की पाण्डुलिपियों में से छाँटकर एकल करना AI से हो सकता है, लेकिन इससे आगे के काम में वर्तमान स्थिति में यह उपयुक्त नहीं होगा।

कुल मिलाकर अभी हमें Documentation के लिए manual काम पर जोड़ देना चाहिए और इसके लिए विशेषज्ञों को तैयार करना होगा।

प्रलेखन(Documentation)

भारत में पाण्डुलिपियों के प्रलेखन का लगभग 250 वर्षों का इतिहास रहा है। मैकेंजी ने सर्वप्रथम इसे प्रस्तुत किया। उन्होंने एक-एक पाण्डुलिपि के लिए उसकी विषयवस्तु का विवरण संक्षेप में संकलित किया। 19वीं शती के आरम्भ से ही दो प्रकार की विवरणी बनती रही।

1. किसी संस्था में संकलित पाण्डुलिपियों के लिए
2. व्यक्तिगत संग्रह की पाण्डुलिपियों के लिए

19वीं शती के अंत से एसियाटिक सोसायटी, कोलकाता, बिहार रिसर्च सोसायटी, पटना, नागरी प्रचारिणी सभा, बनारस, बिहार राष्ट्रभाषा परिषद्, पटना के द्वारा अतीत में सर्वेक्षित पाण्डुलिपियों का प्रलेखन किया गया है। राजा राजेन्द्र लाल मित्र, म.म. हरप्रसाद शास्त्री, सी. सी. बेंडॉल, डा. पीताम्बर दत्त बड़वाल, डा. के.पी. जायसवाल, शिवपूजन सहाय, डा. बेचन झा, डा. अनन्त प्रसाद बनर्जी आदि के निर्देशन में घर-घर सर्वेक्षक को भेज कर सामग्री संकलित की गयी तथा उन्हें व्यवस्थित कर विवरणात्मक हस्तलेखसूची का प्रकाशन कराया गया।

ये प्रकाशित ग्रन्थ आज भी उस पाण्डुलिपि के विषय में पूर्ण सूचनाएँ देते हैं। यहाँ पर डा. अनन्त प्रसाद

बनर्जी द्वारा सम्पादित A Descriptive Catalogue Of Manuscripts In Mithila Vol. 1 से एक पाण्डुलिपि का प्रलेखन उद्धृत किया जा रहा है सन् 1927 ई. में इसका प्रकाशन पटना से हुआ था। इसमें सर्वेक्षण कर व्यक्तिगत संग्रह में स्थित पाण्डुलिपि का विवरण इस प्रकार दिया गया है-

(३१६)

No. 277. प्रायश्चित्तनिरूपणम् By भवदेव । Prāyaścittanirūpaṇam.

Substance:—Character Maithilī. Dark-white paper. Lines 11 on a page. Letters 59 in a line. Folia 51. 12×3½ inches. Appearance, old. Verse and prose. Complete. Not correct. Date, Śaka 1642. Worn-out. Torn. Soiled. Place of deposit, Pandit Śrīkṣant Miśra, Salampur, Ghataho P. O., Darbhanga. A treatise on Dharmaśāstra, dealing with Prāyaścitta by Bhavadeva.

Beginning. ओं नमः ॥

अनादिभवंसंसारपापप्रशमकारणम् ।
स्मरणं बाहुदेवस्य यस्यनौमितमीश्वरम् ॥
मन्वादिस्मृतिमालोक्य सुविचिन्त्ययथाक्रमम् ।
क्रियते भवदेवेन प्रायश्चित्तनिरूपणम् ॥

End.

यथा ह्ययमः ॥
गवामाहार निर्युक्तैर्यवेः शकून्समश्नुतः ।
यान्यं कृच्छ्रं स्मृष्टिं मासेन श्रुतिपारगैः ।
सर्वकर्मपदंशैतत्सर्वकल्मषनाशनम् ।
तत्रापि पञ्चधेनवः ॥
इति कृच्छ्रपरिच्छेदः ॥ ० ॥

Colophon.

इति बाडगडमीभुजङ्गा परनाम्ना भट्टश्रीभवदेवरुतौ प्रायश्चित्तप्रकरणं समाप्तं
मिति ॥ शुभमस्तु ॥ ० ॥
शाके पक्षयुगलं चन्द्रगुणितं श्रीविक्रमस्यामले ।
वर्षमासितितरेणमिलिते पक्षेण भाद्रशुभे ॥
कृष्णस्यायु विना ? प्रनस्य जननं यस्याम भूङ्क्षीयते ।
स्तस्या मेवातिथौ दिनेशशशो पुस्तलि लेखाग्रणीः ॥

विषयः । पञ्चमहापातकमन्त्रपुरीषगोवधेयद्रव्यादिसकलप्रायश्चित्तविचारः ।

Previous notice:—C.C. Pt. I. p. 363; Pt. III. p. 77. R. M. Pt. 9. p. 214. H. S. Vol. II. p. 179. (1915)

यहाँ प्रति पंक्ति अक्षर संख्या, पत्र संख्या तथा प्रति पृष्ठ पंक्ति संख्या- ये तीनों हैं। यहाँ पुष्पिका में उक्त 'शाके पक्षयुगलं चन्द्रगुणिते' शब्द से गणना कर ऊपर 'शाके 1642' दे दिया गया है। साथ

ही, यहाँ विवरणी में previous notice के अन्तर्गत निम्नलिखित पूर्वप्रकाशित विवरणी के सन्दर्भ दिए गये हैं।

C. A. = Catalogue of Mss. in the Asiatic Society of Bengal.

C. C. = Catalogus Catalogorum (Aufrecht).

C.I.O.= Catalogue of Mss. in the India Office Library.

H.P.S.= Catalogue of Mss. edited by Mm. Haraprasad Shastri.

R.M. = Catalog of Mss. edited by Rajendra L.al Mitra.

म.म हरप्रसाद शास्त्री के द्वारा व्यक्तिगत संग्रह की पाण्डुलिपियों से इस प्रकार प्रलेखन किया गया है। Notices Of Sanskrit Manuscripts Vol. 7. से एक पाण्डुलिपि का प्रलेखन इस प्रकार है। इसका प्रकाशन 1871-1891 के बीच में हुआ है।

No. 2398. पूजारत्नाकरः। Substance, country-made paper, 12 × 5 inches. Folia, 142. Lines, 10 on a page. Extent, 2,732 ślokas. Character, Nāgara. Date, ? Place of deposit, Dhamdaha, Zillā Pūrṇiyā, Blaiyālā Jhā. Appearance, fresh. Prose and verse. Generally correct.

Pūjā-ratnākara. Rituals for worshipping various gods and goddesses according to Tāntric rules. By Chandeś'vara. He was minister for Peace and War to a king of Mithilā.

Beginning प्रातः समेत रश्मिषो जलभरैर्नित्यं बोधनेनाहितः

पान्थेत्या वितभूतिचन्दमचयेरासिप्रभाबोष्णः ।

देवदन्द्रकलाविलासितलक्ष्मीरौविवाहोत्सवा-

रक्षे शैलकलाचैव शिखरतामस्योदरः पातु यः ॥

यत्पूजाकुसुमैः शुभैः सुमनसोनेष्कन्ति कल्पद्रुमान्

यस्मिन्नेवमिदं नैरपि सुधां भीक्षुं यतने + तं ।

स श्रीमान् मिथिलावनौन्द्रमहितो मन्त्रीन्द्रचक्षुरः

सद्भाषैः परिभाषितं वितन्ते पूजासु रत्नाकरं ॥

साधारणविधेयाश्चरहः परिकीर्तितः ।

सूर्यपूजातरङ्गोऽयं द्वितीयोऽयं सुविस्तरः ॥

— — — — — चित्रं हि

मानुषे लोके विदिर्भवति कर्मणां ॥

कर्मणां विदिं कां कनोदेवतायां पूजां कुर्यादित्यर्थः । इत्यादि ।

End. त्रितिरिह जपेन्मन्त्री मातरीलोकमातरः ।

कोलिकेन विधानेन रात्र्यायुःसुखदायिकाः ॥

Colophon. इति महासाधुविषयविकसन्नेक्षुरशौचच्छेरविरचितः पूजारत्नाकरः

समाप्तः शुभमस्तु ।

विषयः । साधारणतोद्देशपूजाविधिकथनं । तत्र कालदेशस्थानादिविवेचनं । मण्डप-
बलिहरणादिविधिकथनं । पुष्पचयनादिविधिः । वेदिधितानादिविधिः । नेत्रेशप्रतिपत्तिवि-
धिः । अथ सूर्य्यपूजाया आवयज्ञकताकथनं । तत्फलदादिकथनं । पूजाधिकारिनिश्चयमादिकथनं ।
सूर्य्यमन्दिरमाज्जनादिकथनं । ब्रह्मस्थानविधिः । पञ्चमयविधिः । सप्तमपूजनयोःफलक-
थनं । सूर्य्यावेदानविधिः । पञ्चोपचारविधिः । मन्त्रपुष्पविवेचनं । धूपदौपत्यवेद्यादिप्रतिपत्ति-
कथनं । वासोऽलङ्कारणादिदानविधिः । अथ सूर्य्यनित्यपूजाविधानं । चक्रादिनिर्माणाविधिक-
थनपूर्वकं तत्पूजादिविधिकथनं । खण्डिलपूजादिविधिः । सूर्य्यमूर्त्तिपूजाविधिः । सौरस्थाना-
दिविधिः । सूर्य्यस्य रथयात्राविधिः । सौरधर्मकथनं । अथ शिवपूजाविधिकथनं । तत्र पार्थि-
वादिस्त्रिपूजाविधिः । ब्रह्मसाधारणादिविधिः । पञ्चोपचारदानादिविधिः । घृताभिषेकमहा-
स्थानादिविधिः । पुष्पादिविवेचनं । शिवत्रौतये दानादिविधिः । घृतकम्बलविधिः । प्रतिकृति-
निवेदनविधिः । महास्थानविधिः । शिवपूजायारादिविधानं । शिवस्य रथयात्राविधिः । रत्नं
ब्रह्मणः पूजाविधिः । तत्र रथयात्रादिविधानञ्च । विष्णुपूजाविधिः । तद्गृहमाज्जनादिवि-
धिः । तदुपचारादिविधिः । तत्र पुष्पधूपदौपत्यवेद्यादिविधानं । शिलापरीक्षादिविधिः । भग-
वदौषादौषितानामपराधादिकथनं । दुर्गापूजादिविधिः । कुमारौपूजादिविधिः । मातृपूज-
नादिविधिः । देवीयामविधिः । मूर्त्तिनिर्माणादिविधिः । देव्याः पविचारोपचयविधिः । कौलभेदे
तस्याः पूजनादिकथनञ्च ।

इसमें हम कुछ विशेष बातें पाते हैं ।

1. ग्रन्थ का परिचय अंग्रेजी में दे दिया गया है ।
2. ग्रन्थ का विस्तार (Extent) दिया गया है । इसमें गद्य तथा पद्य दोनों हैं, इसका अर्थ है कि यहाँ उक्त श्लोक संख्या वास्तव में पाण्डुलिपि शास्त्रीय 'ग्रन्थसंख्या' है, जो पत्रसंख्या×पंक्तिसंख्या×अक्षर संख्या को 32 से भाग देने पर प्राप्त होता है । इस प्रकार इस विवरणी में भी प्रतिपंक्ति अक्षर संख्या को गिना गया है ।
3. इसमें विषय विवरण सम्पूर्ण है । पाण्डुलिपिक भीतर के सभी विषयों को पढ़कर संकलित कर लिया गया है ।
4. आरम्भ के अंश को ग्रन्थ परिचय तथा ग्रन्थकार परिचय तक के सभी श्लोकों का संकलन कर लिया गया है ।
5. End text और Colophon में स्पष्ट अंतर यह है कि End text हमेशा ग्रन्थ का भाग होता है तथा ग्रन्थ सम्बन्धी परिचयात्मक अंश Colophon कहलाता है ।

जब हम व्यक्तिगत संग्रह की पाण्डुलिपि की बात करते हैं तो हमें यह मान कर चलना चाहिए कि अगली बार वह पाण्डुलिपि वहाँ नहीं भी मिल सकती है, अतः जितनी देर तक वह प्रलेखक के सामने है, उतने समय में यथासाध्य या तो उसका डिजिटाइजेशन कर लिया जाए या अधिक से अधिक संगत अंश की प्रतिलिपि कर ली जाए ।

प्रस्तावित प्रलेखन की प्रविष्टियाँ

इस प्रकार, निम्नलिखित प्रविष्टियाँ प्रलेखन के लिए उपयुक्त हैं। इससे पाण्डुलिपि का पूर्ण परिचय मिल जाएगा-

1. Institution Name: Person/Institute
2. Address: Full address with contact No. and Email (if any)
3. Manuscript ID
4. Accession No./Bundle No.
5. Previous notice: यदि किसी संस्थान की पाण्डुलिपि है और उस संस्थान से पूर्व में कोई विरणी प्रकाशित हो चुकी है तो उसके सन्दर्भ का उल्लेख होगा।
6. Title: पाण्डुलिपि में जो शीर्षक है।
7. Author: लेखक या संपादक
8. Subject: विषयों की सूची अलग से तैयार करना आवश्यक है। उन्हीं विषयों में से एक को चुनना होगा। वर्तमान में विषय में सर्वेक्षकों ने मनमानी की है। ज्योतिष शास्त्र को ही चार-चार प्रकार से लिखे गये हैं और एक इस्लाम-ज्योतिष अलग है। अतः एक समिति बनाकर विषयों को सूची निर्धारित कर लेनी होगी।
9. Scribe: लिपिकार का नाम, यदि पिता के नाम सहित है तो पिता का भी नाम होगा।
10. Geographical Location of Scribe: जिस स्थान पर पाण्डुलिपि लिखी गयी है, उस स्थान का नाम यहाँ अपेक्षित होगा।
11. Purpose of the scribe: अपने लिए लेखन, दूसरे के लिए लेखन या बिक्री के लिए लेखन का विवरण। यदि लिपिकार ने अंत में यादृश पुस्तकं दृष्ट्वा आदि अस्वीकरण (disclaimer) का व्यवहार किया है तो विक्रयार्थ लेखन माना जा सकता है। स्वार्थ लेखन में अशुद्धि नहीं रहती है। परार्थ लेखन में अक्षर सुन्दर रहते हैं तथा अशुद्धि न्यूनतम होती है। किन्तु वित्क्रयार्थ पाण्डुलिपि में अक्षर तो सुन्दर रहता है किन्तु अशुद्धियों की पर्याप्त सम्भावना रहती है। अतः इस प्रविष्टि से सम्पादन हेतु उत्कृष्ट पाण्डुलिपि की सूचना मिलेगी।
12. Script: लिपि की भी सूची अलगसे निर्धारित करनी होगी। कुछ लिपियाँ नहीं हैं।
13. Language: भाषा की भी सूची निर्धारित कर लेनी चाहिए।
14. Status: Complete / Incomplete:
15. No. of missing folios/ Pages:
16. Material:
17. Condition:
18. Size: यह माप हमेशा सेंटीमीटर में होना चाहिए।
19. Written space: यह माप हमेशा सेंटीमीटर में होना चाहिए।
20. Folios / Pages: यदि पाण्डुलिपि के साथ सादा पृष्ठ अतिरिक्त है तो उसकी संख्या अलग से लिखनी चाहिए। कभी कभी दूसरे ग्रन्थ का अनाथ पृष्ठ (Orphan page) भी साथ पाये जाते हैं तो उसका भी उल्लेख होना चाहिए।
21. No. of lines per page: कम से कम पाँच पृष्ठों पर पंक्ति संख्या गिनकर लिखा जाना चाहिए।
22. No. of letters per line: बीच-बीच से पाँच पंक्तियों में अक्षर संख्या गिनने पर उसका उल्लेख जैसे 35-38, दिया जाएगा।
23. Commentary (if any):
24. Commentator:

25. Language and Script of the Commentary:
 26. No. of Illustrations:
 27. Date of the Manuscript: पाण्डुलिपि में अंकित तिथि को अंतरराष्ट्रीय वर्ष (CE) में बदल कर।
 28. Beginning Line: जहाँ तक ग्रन्थ परिचय है, वहाँ तक लिखना चाहिए।
 29. Ending Line: ग्रन्थ का अन्तिम अंश। कम से कम एक श्लोक या एक पंक्ति पूर्ण
 30. Colophon: पुष्पिका एवं पुष्पिकोत्तर अभिलेख
 31. Publication Status: ग्रन्थ पूर्व प्रकाशित है या नहीं इसका विवरण।
 32. Description: ग्रन्थ का संक्षिप्त विवरण पाण्डुलिपि में उल्लिखित शीर्षक से भिन्न यदि कोई हो तो यहाँ उल्लेख हो सकता है।
 33. First and Last two content folios (Image Uploading):
 34. Date of the Documentation: प्रलेखन की तिथि
 35. Name of Documenter: इसी स्थान पर पाण्डुलिपि संसाधन केन्द्र का नाम दिया जाए।
- इस प्रकार एक फार्म तैयार कर लेना चाहिए, जिसमें यहाँ दिया गया क्रम भी इसी प्रकार रहना चाहिए। इस क्रम में प्रलेखक को समय की बचत होगी तथा बार-बार पाण्डुलिपि को पलटना नहीं पड़ेगा।
- इस प्रकार, पाण्डुलिपि के सर्वेक्षण तथा विवरणात्मक सूची निर्माण के अपने अनुभवों के आधार पर यथामति मैंने सर्वेक्षण तथा प्रलेखन पर अपना प्रस्ताव दिया है।

A Metadata Scheme for Gyan Bharatam with an Emphasis on Relations and AI

Dr. ARD Prasad, Former Professor DRTC (ISI), Bangalore

Abstract

This script discusses the challenges associated with Indian manuscripts and examines the role of metadata in addressing them. It explains the concept of metadata, highlights its importance, and classifies metadata elements into distinct categories. The study presents the metadata elements adopted by Krithi Sampada and evaluates them in comparison with DCMI, TEI, and CIDOC-CRM. It identifies the strengths, redundancies, and missing elements in the NMM metadata scheme, while emphasizing the need for a wide range of relational elements to capture the complex interconnections among manuscripts. To promote standardization, the paper lists relevant authority files and suggests the development of ontologies using identifiers such as titles, author authorities (VIAF), and persistent identifiers like DOI or CNRI handles. The central focus is on compiling a comprehensive set of relationships that can support the formulation of powerful search queries. Furthermore, the paper explores the application of Artificial Intelligence tools for enhancing metadata creation — specifically in data entry, data cleaning, identifying inter-record relations within the repository, and linking with external online resources. It also outlines the types of natural language queries that can be applied to a manuscript repository. Finally, the script proposes a brief framework for selecting suitable software tools and metadata standards to adopt.

Discussion Topics

- Complexity and Challenges Posed by Indian Manuscripts:
Diversity of Scripts and Languages; Paleographic Challenges; Diglossia in Texts; Materials and Formats; Complexity of Documentation; Diverse Genres; Cultural Significance; Oral-Written Interface; Geographic Distribution; Colonial Records; Ownership and Access; Specialized Skills
- Importance of Metadata in Gyan Bharatam
Importance of Metadata in the Context of Manuscripts: Discovery and Accessibility; Preservation of Contextual Information; Interoperability and Standardization; Support for Scholarly Research; Cultural Sensitivity and Ethical Management;
Advantages of Metadata in the Context of Manuscripts: Enhanced

Digital Access; Improved Preservation; Facilitation of Advanced Technologies; Support for Paleographic and Codicological Analysis; Cultural and Historical Insights; Streamlined Collaboration; Resolution of Ownership and Access Issues; Support for Training and Education.

- Categories of Metadata Elements

Descriptive Metadata; Structural Metadata; Administrative Metadata; Technical Metadata; Preservation Metadata; Relational Metadata; Linking Metadata; Cultural and Contextual Metadata; Rights and Legal Metadata;

- Overview of NMM's Metadata Schema

An Evaluation of NMM Metadata Elements; List of NMM Metadata Elements; Strengths of the NMM Schema; Some Missing Elements; Redundant Elements; Recommendations

- Authority Files Required for Manuscript Metadata

Personal Name Authority File; Title Authority File; Subject/Genre Authority File; Language Authority File; Script Authority File; Geographic/Place Authority File; Organization/Institution Authority File; Material Authority File; Personal Name Authority File; Catalogue Authority File; Implementation in Repositories

- Bibliographic Relations

- Work-Level Relations: hasWork/ isWorkOf; derivedFrom/ hasDerivation; translationOf/ hasTranslation; commentaryOn/ hasCommentary; abridgmentOf/ hasAbridgment; paraphraseOf/ hasParaphrase
- Manuscript-to-Manuscript Relations: copyOf/ hasCopy; exemplarOf/ hasExemplar; fragmentOf/ hasFragment; boundWith/ isBoundWith;
- Collection Relations: partOf/ hasPart; seriesOf/ hasSeries;
- Intertextual Relations: quotesFrom/ isQuotedIn; alludesTo/ isAlludedToIn; respondsTo/ hasResponse; refutesFrom/ isRefutedIn;
- Structural Relations: precededBy/ follows; illustratedBy/ illustrates; glossedBy/ glosses;
- Authorship and Attribution Relations: authoredBy/ isAuthorOf;

- attributedTo/ isAttributedWork; compiledBy/ isCompilationOf;
translatedBy/ isTranslationOf;
- Scribal Relations: scribedBy/ isScribeWork; copiedBy/
isCopyOf; correctedBy/ isCorrectionOf;
 - Patronage Relations: patronizedBy/ isPatronageOf;
commissionedBy/ isCommissionOf; dedicatedTo/
isDedicationOf;
 - Temporal Relations: contemporaryWith/ isContemporaryOf;
dynastyOf/ hasDynasty;
 - Ownership Relations: ownedBy/ isOwnerOf; inheritedBy/
isInheritanceOf; donatedTo/ isDonationOf;
 - Geographical Relations: Spatial Relations: originatesFrom/
isOriginOf; producedIn/ isProductionSiteOf; foundIn/
isFindSiteOf;
 - Cultural Geographic Relations: regionOf/ hasRegion; cultureOf/
hasCulture;
 - Language Relations: writtenIn/ isLanguageOf; bilingualWith/
isBilingualOf; dialectOf/ hasDialect;
 - Script Relations: scriptOf/ hasScript; paleographyOf/
hasPaleography;
 - Subject Relations: subjectOf/ hasSubject; isciplineOf/
hasDiscipline; traditionOf/ hasTradition; disciplineOf/
hasDiscipline: philosophyOf/ hasPhilosophy; theologyOf/
hasTheology;
 - Material Relations: madeFrom/ isMaterialOf; inkOf/ hasInk;
bindingOf/ hasBinding;
 - Technical Relations: techniqueOf/ hasTechnique; processOf/
hasProcess;
 - Preservation and Conservation Relations: Condition Relations:
damagedBy/ isDamageSource; treatedBy/ isTreatmentOf;
digitizedBy/ isDigitizationOf;
 - Scholarly and Research Relations: studiedBy/ isStudyOf;
catalogedBy/ isCatalogOf; publishedBy/ isPublicationOf;

- Comparative Relations: comparedWith/ isComparisonOf;
- Cermonial and Ritual Relations: Functional Relations: usedFor/ isFunctionOf; performedIn/ isPerformanceVenue; liturgyOf/ hasLiturgy;
- Sacred Relations: consecratedFor/ isConsecrationOf; blessedBy/ isBlessingOf;
- Probable Role of Artificial Intelligence in Metadata
 - Generating Metadata from Unstructured Data
 - AI-based Text Summarisation
 - Querying the Database using Natural Language
 - How AI Tools Assist in Data Entry for Relations
 - AI Capabilities for External Resource Integration
 - Ensuring Error-Free metadata
 - Criteria for Metadata Cleaning Tools
 - A List of Open-Source Metadata Cleaning Tools
- Conclusion
- References
- Epilogue on Koba Manuscript Metadata Search Portal

The National Mission for Manuscripts (NMM), established in 2003 by the Ministry of Culture, Government of India, developed a metadata schema to document and preserve India's vast manuscript heritage, estimated at ten million manuscripts. This schema, based on the Dublin Core Metadata Initiative (DCMI), is designed to comprehensively catalogue manuscripts, including those in Indic scripts such as Devanagari, Tamil, and others, and is used in the NMM's Kriti Sampada National Database of Manuscripts.

1. Complexity and Challenges Posed by Indian Manuscripts

The vast number of ancient and medieval manuscripts in various languages and scripts poses challenges in organising them for retrieval purposes. The challenges, in a way, guide the system development in choosing metadata elements and thereby affect the retrieval.

Diversity of Scripts and Languages

- Indian manuscripts encompass an extraordinary variety of scripts and languages, including widely used ones like Devanagari, Tamil, Bengali, and Telugu, as well as archaic scripts such as Brahmi, Kharosthi, and Grantha. These scripts represent a linguistic heritage spanning millennia, with some, like Sharada or Modi, now rarely used but critical for historical research. The collection comprises texts in Sanskrit, Pali, Prakrit, Tamil, and other Indian languages, as well as Persian and Arabic manuscripts from medieval India.

Paleographic Challenges

- The same script can exhibit significant variations across different time periods, regions, and scribal traditions, posing challenges for accurate identification and interpretation. For example, Devanagari scripts from 10th-century North India differ markedly from those in 18th-century Maharashtra. Multispectral imaging and high-resolution digitization further aid in analysing scripts that have faded or been altered over time, ensuring precise documentation and study.

Diglossia in Texts

- Many Indian manuscripts, such as the plays of Kalidasa, reflect diglossia, where distinct linguistic registers are used based on social hierarchy, gender, or context. In Kalidasa's works, for instance, male characters of high status often speak in Sanskrit, while female characters or those of lower status use Prakrit, reflecting a deliberate linguistic stratification.

Materials and Formats

- Indian manuscripts were inscribed on diverse materials, including palm leaves, birch bark, handmade paper, cloth, and occasionally metal plates, each requiring specialized preservation techniques. Palm-leaf manuscripts, often long and narrow, were tied with cords, while paper codices and scrolls reflect later formats. Advanced imaging technologies, such as 3D scanning for textured surfaces, ensure that physical characteristics like binding techniques or material wear are documented alongside textual content.

Complexity of Documentation

- Documenting Indian manuscripts requires meticulous care due to

their fragility and historical significance. This calls for specialized imaging techniques, such as multispectral and hyperspectral imaging, to recover faded or illegible texts without physical handling. These technologies support both preservation and global access, allowing scholars to study manuscripts remotely without risking damage to originals.

Diverse Genres

- The collection spans a wide range of genres, from sacred Vedic hymns and Puranic narratives to scientific treatises like Aryabhata's mathematical and astronomical works, medical texts like the Charaka Samhita (Ayurveda), and poetic masterpieces by poets like Kalidasa and Bhavabhuti. The digital platform requires organizing these genres into searchable categories, with curated exhibits that highlight their historical and intellectual significance.

Cultural Significance

- Many manuscripts hold profound ritual, religious, or cultural value, such as Jain cosmological texts or Tantric ritual manuals, necessitating sensitive handling and ethical access policies. The Gyan Bharatam should collaborate with cultural custodians, including temple authorities and community leaders, to ensure respectful management. Digital access is designed with cultural sensitivities in mind, with restricted sections for sacred texts and consultation processes to honour traditional protocols, ensuring that digitization respects the manuscripts' spiritual roles.

Oral-Written Interface

- Indian manuscripts often served as mnemonic aids for oral traditions rather than standalone records, particularly in Vedic and epic literature. For example, the Rigveda was primarily transmitted orally, with manuscripts acting as references for reciters. The digital platform includes audio archives of oral recitations, linked to corresponding manuscript texts, to preserve this interplay.

Geographic Distribution

- Indian manuscripts are dispersed across diverse locations, from major institutions like the National Manuscript Mission in Delhi to private collections, temples, and monasteries in remote areas like Kerala's villages or Himalayan monasteries. The Gyan Bhartam should

partner with these entities to create a centralized digital repository, using secure cloud storage and standardised metadata to unify access.

Colonial Records

- British colonial scholars, such as those at the Asiatic Society, documented many Indian manuscripts but often introduced errors due to their limited familiarity with Indian languages, scripts, or cultural contexts. These records frequently lack detail or misinterpret texts, complicating modern scholarship. The digital archive should include tools to cross-reference colonial records with primary sources. Scholarly annotations provide context on colonial influences, helping researchers navigate these historical inaccuracies.

Ownership and Access

- Manuscripts in private or religious collections often face issues of contested ownership or restricted access due to cultural, religious, or familial sensitivities, warranting the establishment of ethical access frameworks, negotiating with custodians to digitize manuscripts while respecting their rights. Blockchain-based access logs ensure transparency in usage, while restricted-access portals protect sensitive materials. Community engagement programs educate stakeholders on the benefits of digitization, fostering trust and collaboration.

Specialized Skills

- The study of Indian manuscripts requires expertise in paleography, codicology, and specific languages or scripts, but such scholars are scarce. Training new experts is a resource-intensive process that requires years of linguistic and historical study. The NMM has been addressing these issues by offering training programs.

2. Importance of Metadata in Gyan Bharatam

Metadata is structured information that describes, explains, locates, or otherwise facilitates the retrieval, use, or management of a resource. Often referred to as “data about data,” it provides essential details about a manuscript, such as its title, author, date, language, script, material, provenance, and content summary, without reproducing the full text. In the context of Indian manuscripts, metadata might include specifics like the script (e.g., Devanagari, Tamil, or Brahmi), the physical condition (e.g., palm leaf or paper, signs of wear), the genre (e.g., Vedic hymn, mathematical treatise),

or its cultural significance (e.g., sacred ritual text). Metadata is typically stored in standardized formats, such as Dublin Core or MARC, to ensure consistency and interoperability across digital systems.

Importance of Metadata in the Context of Manuscripts

Metadata is critical for managing, preserving, and accessing Indian manuscripts, which are often ancient, fragile, and written in diverse scripts and languages. Its importance lies in addressing the unique challenges of these manuscripts, such as their scattered distribution, varied formats, and cultural significance. Below are key reasons why metadata is essential:

Discovery and Accessibility

- Indian manuscripts are dispersed across institutions, temples, private collections, and remote locations like Kerala villages or Himalayan monasteries. Metadata provides a standardized way to catalogue these manuscripts, making them discoverable through digital databases. For example, a scholar searching for Tamil palm-leaf manuscripts can use metadata fields like “script: Tamil” or “genre: Saiva poetry” to locate relevant items without physically accessing them.
- Metadata enables global access by integrating manuscripts into centralized repositories, such as the National Manuscript Mission’s database, allowing researchers worldwide to study texts without travel.

Preservation of Contextual Information

- Metadata captures critical details about a manuscript’s physical and cultural context, such as its material (e.g., birch bark, palm leaf), dimensions, condition, and provenance. This is vital for fragile manuscripts, where physical handling must be minimized. For instance, metadata noting “palm leaf, 12th century, insect damage” informs conservators about preservation needs.
- It preserves historical and cultural context, such as whether a manuscript is a sacred Jain text or a secular astronomical treatise by Aryabhata, ensuring that its significance is not lost during digitization.

Interoperability and Standardization

- With the diversity of Indian scripts (e.g., Brahmi, Grantha, Sharada) and languages (e.g., Sanskrit, Prakrit, Tamil), metadata ensures consistency by using standardized formats. This allows different

institutions to share and compare data seamlessly, creating unified catalogues across regions like Delhi, Pune, or Tamil Nadu.

- Standardized metadata supports integration with global platforms, enabling Indian manuscripts to be part of any international archives like WorldCat or digital humanities projects.

Support for Scholarly Research

- Metadata enables precise searches for specific genres, authors, or historical periods, facilitating research on topics like Vedic literature or Ayurvedic texts. For example, a scholar studying Kalidasa's plays can filter manuscripts by "author: Kalidasa" and "language: Sanskrit/Prakrit" to analyse diglossic patterns.
- It aids paleographic and codicological studies by documenting script variations, scribal hands, or material characteristics, helping scholars date manuscripts or trace regional differences.

Cultural Sensitivity and Ethical Management

- Many Indian manuscripts have ritual or religious significance, such as Tantric texts or Vedic hymns, and metadata can include access restrictions to respect cultural sensitivities. For instance, metadata might flag a manuscript as "restricted: temple use only," ensuring ethical handling.
- It tracks ownership and provenance, addressing disputes over manuscripts in private or religious collections, which is common in India due to colonial legacies or family archives.

Advantages of Metadata in the Context of Manuscripts

The use of metadata in managing Indian manuscripts offers several practical advantages, enhancing their preservation, study, and accessibility while addressing their unique challenges.

Enhanced Digital Access:

- Metadata enables the creation of searchable digital archives, allowing users to access manuscripts remotely. For example, a researcher in Europe can study a Tamil manuscript in a Kerala temple through a digital platform, using metadata filters like "language: Tamil" or "material: palm leaf."
- It supports multilingual interfaces, accommodating India's linguistic

diversity (e.g., Sanskrit, Tamil, Persian) by including metadata in multiple languages or scripts, broadening access for non-specialists.

Improved Preservation:

- By documenting physical characteristics (e.g., “birch bark, brittle condition”), metadata guides conservation efforts, prioritizing manuscripts for urgent digitization or repair. This is critical for fragile materials prone to decay in India’s humid climate.
- Digital metadata reduces the need for physical handling, as scholars can access detailed descriptions and high-resolution images instead of the original manuscript, minimizing wear and tear.

Facilitation of Advanced Technologies:

- Metadata supports AI-driven tools, such as OCR for recognizing scripts like Devanagari or Grantha, by providing training data on script types and languages. For instance, metadata noting “script: Kharosthi” helps AI models identify and transcribe ancient texts.
- It enables multispectral imaging workflows by cataloging manuscripts that require specialized imaging for faded or damaged text, ensuring efficient resource allocation.

Support for Paleographic and Codicological Analysis:

- Metadata captures variations in scripts (e.g., 10th-century vs. 18th-century Devanagari) and material formats (e.g., palm-leaf folios vs. paper codices), aiding paleographers in dating manuscripts or tracing scribal traditions.
- It supports codicological studies by recording binding techniques, folio counts, or marginal notes, providing insights into manuscript production and use.

Cultural and Historical Insights:

- Metadata documenting a manuscript’s genre (e.g., Purana, mathematical treatise) or cultural role (e.g., ritual text) helps scholars understand its historical context. For example, metadata linking a manuscript to a specific monastery can reveal its role in Buddhist education.
- It preserves the oral-written interface by noting if a manuscript, like a

Vedic text, was used as a mnemonic aid, linking it to audio archives of recitations for a holistic study.

Streamlined Collaboration:

- Metadata enables institutions, such as the Bhandarkar Oriental Research Institute or local temples, to share data through standardized formats, fostering collaborative preservation efforts across India's diverse regions.
- It supports partnerships with global archives, allowing Indian manuscripts to be studied alongside other world heritage texts, enhancing their global visibility.

Resolution of Ownership and Access Issues:

- Metadata tracks provenance and ownership, using blockchain-based systems to ensure transparency in contested cases, such as manuscripts in private collections or colonial archives.
- It facilitates ethical access by flagging culturally sensitive manuscripts, allowing custodians to set restrictions while still enabling scholarly study through controlled digital access.

Support for Training and Education:

- Metadata databases serve as educational tools, providing scholars and students with detailed information on scripts, languages, and genres, reducing the dependency on scarce paleographic expertise.
- It supports the creation of online training modules, where metadata-driven examples (e.g., images of Brahmi script variations) help train new scholars in manuscript studies.

Metadata is the backbone of managing Indian manuscripts, transforming them from scattered, fragile artifacts into accessible, preserved, and researchable resources. By providing structured information about their content, physical characteristics, and cultural context, metadata ensures that these manuscripts—ranging from Vedic hymns to Ayurvedic texts—are discoverable, preserved, and studied ethically. Its advantages include enhanced digital access, improved preservation, support for advanced technologies, and streamlined collaboration, all of which address the challenges of India's diverse scripts, languages, and cultural sensitivities. Through metadata, the rich heritage of Indian manuscripts can be safeguarded and shared with global audiences while respecting their historical and cultural significance.

3. Categories of Metadata Elements

Descriptive Metadata

- Purpose: Describes the intellectual content and identity of the manuscript to aid discovery and interpretation.
- Examples:
 - Title: The primary title of the manuscript (e.g., “Bhagavad Gita”).
 - Other Title: Alternative or variant titles (e.g., “Gita”).
 - Author: The creator of the text (e.g., “Vyasa”).
 - Commentator: Author of any commentary (e.g., “Adi Shankara” for “Sankara Bhashya”).
 - Subject: Thematic content (e.g., “Yoga Philosophy,” “Ayurveda”).
 - Keywords: Tags for searchability (e.g., “Vedic,” “Jainism”).
 - Language: Primary and secondary languages (e.g., “Sanskrit,” “Tamil commentary”).
 - Script: Writing system (e.g., “Devanagari,” “Grantha”).
 - Genre: Type of text (e.g., “Treatise,” “Poetry,” “Ritual Manual”).
 - Abstract/Summary: Brief content description.
- Relevance to Manuscripts: These elements enable precise searching and contextual understanding, critical for Indian manuscripts with diverse languages and subjects. For example, NMM’s Manus Granthavali includes “Subject” and “Commentary” to link texts like a Purana with its regional interpretations.

Structural Metadata

- Purpose: Describes the physical and logical structure of the manuscript, including its components and organization.
- Examples:
 - Folio Number: Total folios or pages (e.g., “50 folios”).
 - Bundle Number: Grouping of folios in a set (e.g., for palm-leaf bundles).
 - Page Sequence: Order of pages or sections, especially for fragmented

manuscripts.

- Illustrations: Presence and type of illustrations (e.g., “Miniature paintings”).
- Marginalia: Notes or annotations in margins.
- Colophon: Information about scribe, date, or place at the manuscript’s end.
- Parts/Sections: Divisions like chapters or sub-texts (e.g., “Sutra” and “Bhashya”).
- Relevance to Manuscripts: Structural metadata preserves the manuscript’s physical integrity in digital form, as seen in NMM’s use of “Folio Number” to track palm-leaf sequences. METS excels here with its <structMap> for linking pages.

Administrative Metadata

- Purpose: Manages ownership, provenance, and access details to support ethical handling and institutional management.
- Examples:
- Manuscript Number: Unique identifier (e.g., “MS-12345”).
- Provenance: Ownership history (e.g., “Acquired from Kerala temple, 1800s”).
- Location: Current repository (e.g., “Oriental Research Institute, Mysore”).
- Access Rights: Restrictions (e.g., “Restricted due to sacred status”).
- Custodian: Current holder, often for private or temple collections.
- Catalogue Source: Reference to printed or external catalogues.
- Acquisition Date: When the manuscript entered the repository.
- Funding Source: For digitization or conservation efforts.
- Relevance to Manuscripts: These elements address ethical concerns like repatriation (e.g., colonial-era manuscripts) and access control for sacred texts, as seen in NMM’s collaboration with local custodians.

Technical Metadata

- Purpose: Documents the technical aspects of digitization and digital preservation to ensure long-term accessibility.
- Examples:
- File Format: Format of digital files (e.g., “TIFF,””PDF”).
- Resolution: Image quality (e.g., “600 DPI”).
- File Size: Storage requirements (e.g., “50 MB”).
- Color Mode: Grayscale, RGB, or multispectral.
- Scanner Make/Model: Equipment used (e.g., “Epson DS-7000”).
- Software: Tools for digitization or metadata creation (e.g., “Adobe Photoshop”).
- Date Created/Modified: Timestamp for digital files.
- Checksum: For data integrity verification.
- Compression: Method used (e.g., “LZW”).
- Relevance to Manuscripts: Critical for fragile Indian manuscripts, where high-resolution imaging (e.g., multispectral for faded texts) is common. NMM’s guidelines emphasize TIFF at 600 DPI for archival quality.

Preservation Metadata

- Purpose: Records condition and conservation details to guide physical and digital preservation.
- Examples:
- Condition: Physical state (e.g., “Fragile, insect damage”).
- Material: Substrate (e.g., “Palm leaf,””Birch bark”).
- Dimensions: Length and width (e.g., “10 x 4 inches”).
- Conservation History: Treatments applied (e.g., “Laminated in 1995”).
- Storage Conditions: Current environment (e.g., “Climate-controlled vault”).
- Preservation Events: Actions like digitization or rebinding.

- Relevance to Manuscripts: Preservation metadata ensures proper handling of delicate materials, as NMM does by noting “Condition” to prioritize conservation.

Relational Metadata

- Purpose: Captures relationships between manuscripts, such as derivations, commentaries, or copies, to reflect intellectual and historical connections.
- Examples:
 - IsCommentaryOf: Links a commentary to its base text (e.g., “Sankara Bhashya” to “Bhagavad Gita”).
 - IsCopyOf: Indicates a manuscript is a copy of another.
 - IsPartOf: Connects a manuscript to a larger work or series.
 - HasTranslation: Links to translated versions (e.g., Pali text to Tamil translation).
 - QuotedIn: References where the text appears in other works.
 - RelatedItem: General relations (e.g., manuscripts by the same scribe).
 - Filiation: TEI-specific element for textual lineage (e.g., parent-child relations between versions).
- Relevance to Manuscripts: Indian manuscripts often have complex intertextual relationships (e.g., Vedic texts with commentaries). TEI’s <relatedItem> and <filiation> elements are ideal, though NMM uses simpler fields like “Commentary” to achieve this.

Linking Metadata

- Purpose: Facilitates connections to internal (within the repository) and external resources (e.g., other databases, publications) for broader access and context.
- Examples:
 - Unique Identifier: Internal link (e.g., NMM’s “Manus Id” like “MG-45678” linking to images).
 - URI/DOI: External links to resources like WorldCat, JSTOR, or other repositories.

- Related Catalogue: References to external catalogues (e.g., “Asiatic Society Catalogue”).
- External Resource URL: Links to scholarly articles or related digital collections.
- API Endpoint: For dynamic linking to other systems.
- Relevance to Manuscripts: Linking metadata creates a networked repository, as seen in NMM’s integration with IGNCA or potential links to global archives holding Indian manuscripts.

Cultural and Contextual Metadata

- Purpose: Captures cultural, historical, or religious significance to respect the manuscript’s context and ethical considerations.
- Examples:
 - Cultural Significance: Role in rituals or communities (e.g., “Used in Jain temple ceremonies”).
 - Historical Context: Period or events related to creation (e.g., “Written during Chola dynasty”).
 - Regional Variant: Specific regional traditions (e.g., “Kashmiri Sharada script”).
 - Ownership Disputes: Notes on contested provenance (e.g., “Taken to British Museum, 1850”).
 - Access Restrictions: Cultural or legal limits (e.g., “Restricted to temple priests”).
 - Relevance to Manuscripts: Essential for Indian manuscripts with sacred or communal value, ensuring ethical handling, as NMM does by noting custodians.

Rights and Legal Metadata

- Purpose: Documents intellectual property, copyright, and usage rights to manage access and reproduction.
- Examples:
 - Copyright Status: Public domain or restricted.
 - License: Terms of use (e.g., “Creative Commons BY-NC”).

- Rights Holder: Institution or community owning rights.
- Publication Status: Whether the manuscript has been published or edited.
- Relevance to Manuscripts: Critical for addressing repatriation or access disputes, especially for manuscripts in global collections.

The categories of metadata elements—descriptive, structural, administrative, technical, preservation, relational, linking, cultural, and rights—form a holistic framework for manuscript repositories. They ensure exhaustive documentation, capture complex inter-manuscript relations, and enable internal/external linking, as demonstrated by NMM’s approach. For Indian manuscripts, these categories address the diversity of scripts, languages, and cultural contexts, making repositories like Kriti Sampada powerful tools for preservation and research.

4. Overview of NMM’s Metadata Schema

The NMM’s metadata schema is tailored for manuscripts, including palm leaf manuscripts, and supports detailed documentation of individual manuscripts and collections. It is based on the Dublin Core Metadata Initiative (DCMI), which provides a set of 15 core elements (e.g., Title, Creator, Subject) to describe resources, but the NMM extends these with additional elements specific to manuscripts, such as physical condition, script, and material. The schema is implemented in the Manus Granthavali software, which facilitates data collection in three formats: Manus Data Sheet, Questionnaire, and CAT-CAT (catalog of catalogs). The schema supports multilingual entries, particularly Indic scripts, using Unicode.

An Evaluation of NMM Metadata Elements

The National Mission for Manuscripts (NMM) metadata schema is based on the Dublin Core Metadata Initiative (DCMI), extended with manuscript-specific elements to document India’s rich manuscript heritage. The schema is implemented in the Manus Granthavali software and used in the Kriti Sampada database, focusing on descriptive, administrative, and physical attributes of manuscripts. From the available sources, the schema includes 24 elements, primarily from the Manus Data Sheet, which captures individual manuscript details. Below, I first list the elements, then critically evaluate the schema, identifying missing elements (compared to standards like TEI for manuscripts or CIDOC-CRM for cultural heritage) and redundant ones (e.g., overlapping or duplicative fields). The evaluation considers the schema’s strengths in handling Indic scripts and manuscript-specific details but highlights gaps in digital preservation, multilingual support, and relational metadata.

List of NMM Metadata Elements

The following 24 elements are derived from the guidelines for digitization & the Manus Data Sheet:

1. Manuscript number: Unique identifier for the manuscript.
2. Title: Primary title of the manuscript.
3. Other title: Alternative or variant titles.
4. Author: Creator or writer of the manuscript.
5. Organization: Institution or entity associated with the manuscript (e.g., repository).
6. Commentary: Textual notes or explanations within or about the manuscript.
7. Commentator: Individual providing commentary.
8. Scribe: Person who transcribed or wrote the manuscript.
9. Language: Language(s) in which the manuscript is written.
10. Script: Writing system used (e.g., Devanagari, Grantha).
11. Complete/Incomplete: Indicates if the manuscript is fully intact.
12. Subject: Thematic content or topic.
13. Bundle number: Group or bundle identifier for manuscripts.
14. Folio number: Numbering of individual folios or pages.
15. Pages: Total number of pages.
16. Material: Physical substrate (e.g., palm leaf, paper).
17. Missing portion: Details on any absent or damaged parts.
18. Illustrations: Presence and description of visual elements.
19. Condition: Physical state (e.g., fragile, damaged).
20. Catalogue source: Reference to the catalog from which the manuscript is documented.
21. Remarks: Additional notes or observations.
22. Manuscript date: Date of creation or copying.
23. Manuscript length (in inches): Physical length measurement.
24. Manuscript width (in inches): Physical width measurement.

Additionally, the schema includes technical metadata (auto-generated for digitized manuscripts), such as file format (e.g., TIFF), resolution (e.g., 300 DPI), and compression, though these are not part of the core subject metadata.

Strengths of the NMM Schema

- Alignment with Dublin Core: It adopts core elements like Title, Author, Language, and Subject, ensuring basic interoperability with global standards.
- Manuscript-Specific Focus: Elements like Material, Script, Condition, and Folio number are tailored to physical manuscripts, supporting Indic scripts (e.g., Devanagari, Grantha) via Unicode, which is crucial for India's heritage.
- Practicality: The schema is simple and extensible, facilitating data entry via the Manus Data Sheet and Questionnaire, with over 5.2 million records created.
- Support for Digitization: Technical metadata (e.g., resolution) aids preservation, aligning with NMM's digitization guidelines (e.g., 300 DPI TIFF for master copies).

Some Missing Elements

The NMM schema is functional for basic manuscript description but lacks depth in areas like digital preservation, semantic relations, accessibility, and provenance, compared to advanced standards like TEI (Text Encoding Initiative) for manuscripts or CIDOC-CRM for cultural heritage. Missing elements include:

Rights and Licensing: No explicit field for copyright status, access rights, or usage restrictions (e.g., public domain for ancient manuscripts). Essential for digital access and could be added as to prevent misuse.

Provenance Details: Lacks fields for ownership history, acquisition date, or custodians (e.g., “inheritedBy” or “donatedTo” from our prior discussions). Important for historical manuscripts.

Digital-Specific Metadata: Beyond basic technical metadata, missing elements like file format, checksum, or digital preservation status (e.g., PREMIS for migration history). Critical for digitized manuscripts; NMM's focus on physical attributes overlooks long-term digital sustainability.

Accessibility Features: No fields for accessibility (e.g., alt-text for illustrations, WCAG compliance). Limits use for diverse users.

Semantic Relations: Elements like Commentary or Commentator could be

relational, but the schema treats them as stand alone. Missing explicit support for relations like hasTranslation, hasWork, or authoredBy.

Cultural and Linguistic Depth: While Script and Language are included, missing fields for dialect, paleography, or cultural significance (e.g., “cultureOf”). Insufficient for India’s multilingual heritage.

Versioning and Updates: No fields for revision history or updates (e.g., digitization date, last modified). Essential for tracking changes; add ``<dc:dateAccepted>``.

Unique Identifiers: Lacks integration with external authority files (e.g., VIAF for creators). Critique: Limits global interoperability; add ``<dc:creator ref=“http://viaf.org/viaf/32307241”>``.

The schema is descriptive but not relational or semantic, limiting advanced queries. It excels in physical description but lacks digital and linked data support, potentially hindering integration with global standards like those in Europeana or WorldCat.

Redundant Elements

Some elements overlap, leading to redundancy, which could be streamlined for efficiency:

Complete/Incomplete and Missing portion: Both describe completeness; Missing portion provides more detail, making Complete/Incomplete somewhat redundant. Critique: Merge into a single “Completeness Status” field with descriptions.

Folio number and Pages: Folio number tracks individual leaves, while Pages counts total; for non-foliated manuscripts, they overlap. Critique: Use one field (e.g., “Page/Folio Count”) with qualifiers.

Manuscript length and Manuscript width: These dimensional measurements are specific but could be combined into “Dimensions” (e.g., “Length x Width in inches”). Critique: Redundant if not always applicable; consolidate for simplicity.

Bundle number and Organization: Bundle number groups manuscripts, while Organization identifies the repository; they overlap in collection-level metadata. Critique: Bundle number could be a sub-field of Organization.

Remarks: A catch-all for additional notes, which may duplicate information in Description or Condition. Critique: Broad and unstructured; limit to non-standardized data to avoid redundancy. Redundancies arise from the schema’s focus on physical manuscripts, leading to overlapping descriptive fields. Streamlining could reduce data entry errors and improve usability,

Recommendations

- Add Relations: Incorporate relational metadata using TEI or CIDOC-CRM (e.g., `<relatedItem type="hasCommentary"/>` for Commentary/Commentator).
- Address Missing Elements: Extend with digital preservation (e.g., PREMIS), accessibility, and provenance fields for modern needs.
- Tools: Use OpenRefine for reconciliation with VIAF and Oxygen XML Editor for TEI extensions.

5. Authority Files Required for Manuscript Metadata

Authority files are standardized datasets or vocabularies that ensure consistency and interoperability in manuscript metadata. They are essential for managing the diversity of Indian manuscripts, enabling accurate documentation, linking related records, and supporting research. Below are the key types of authority files required, with examples and their relevance to manuscript repositories.

Personal Name Authority File

- Purpose: Standardizes names of authors, commentators, scribes, and other individuals associated with manuscripts to resolve variations and ambiguities.
- Examples:
 - Normalized Name: Vyasa (for author of the Mahabharata).
 - Variant Names: Veda Vyasa, Krishna Dvaipayana.
 - Unique Identifier: E.g., VIAF ID (Virtual International Authority File, e.g., VIAF:31569536 for Kalidasa).
 - Role: Author, Commentator, Scribe, Patron.
 - Cultural Context: E.g., “Adi Shankara” linked to Advaita Vedanta tradition.
- Relevance to Manuscripts: Indian manuscripts often list authors or commentators with variant spellings (e.g., “Sankara” vs. “Shankaracharya”) or regional names. An authority file ensures a single canonical form, linking to related manuscripts (e.g., all commentaries by Shankara). NMM’s Manus Granthavali uses a basic name authority to normalize entries.

- External Links: Connects to global standards like VIAF or ISNI (International Standard Name Identifier) for interoperability with repositories like the British Library.

Title Authority File

- Purpose: Standardizes manuscript titles to account for variations, translations, or alternative names, facilitating search and relation mapping.
- Examples:
 - Canonical Title: Bhagavad Gita.
 - Variant Titles: Gita, Srimad Bhagavad Gita, Krishna Gita.
 - Unique Identifier: E.g., Uniform Title in MARC (Machine-Readable Cataloging) or a local ID like NMM's Manus Id.
 - Related Titles: Links to commentaries (e.g., "Sankara Bhashya") or translations.
- Relevance to Manuscripts: Indian manuscripts often have multiple titles (e.g., "Yoga Sutras" vs. "Patanjali Sutras") or regional variants. This file ensures consistent cataloging and supports relational metadata (e.g., linking a base text to its commentaries). NMM uses title fields in its metadata scheme to track such relations.
- External Links: Can reference external catalogues like Aufrecht's Catalogus Catalogorum for cross-repository consistency.

Subject/Genre Authority File

- Purpose: Provides a controlled vocabulary for subjects and genres to categorize manuscripts thematically and support discovery.
- Examples:
 - Controlled Terms: Ayurveda, Vedanta, Bhakti Poetry, Astronomy, Tantra.
 - Hierarchical Structure: E.g., "Philosophy > Vedanta > Advaita."
 - Standard: Library of Congress Subject Headings (LCSH) or a custom thesaurus (e.g., NMM's subject list).
 - Keywords: Specific terms like "Charaka Samhita" for medical texts.

- **Relevance to Manuscripts:** Indian manuscripts cover diverse disciplines (e.g., Vedic hymns, mathematical treatises). A subject authority ensures consistent classification, enabling queries like “all manuscripts on Jainism.” NMM’s scheme includes a “Subject” field for this purpose.
- **External Links:** Maps to LCSH or UNESCO Thesaurus for global interoperability.

Language Authority File

- **Purpose:** Standardizes language names and codes to ensure consistent identification of manuscript languages.
- **Examples:**
 - **Canonical Name:** Sanskrit, Tamil, Pali.
 - **Language Code:** ISO 639-2 (e.g., “san” for Sanskrit, “tam” for Tamil).
 - **Variant Names:** E.g., “Prakrit” vs. regional dialects like Magadhi.
- **Relevance to Manuscripts:** Indian manuscripts use numerous languages, including extinct ones like Prakrit. An authority file resolves ambiguities (e.g., “Tamil” vs. “Dravidian Tamil”) and supports multilingual searches. NMM’s “Language” field relies on such standardization.
- **External Links:** Aligns with ISO 639 standards for integration with global linguistic databases.

Script Authority File

- **Purpose:** Standardizes script names to account for variations and historical changes in writing systems.
- **Examples:**
 - **Canonical Script:** Devanagari, Grantha, Sharada.
 - **Variant Names:** E.g., “Nagari” for Devanagari.
 - **Historical Context:** E.g., “Brahmi, 3rd century BCE.”
- **Relevance to Manuscripts:** Indian manuscripts use diverse scripts, some obsolete (e.g., Kharosthi). This file ensures accurate identification, critical for paleographic studies, and supports relational metadata by linking manuscripts in the same script. NMM’s “Script” field uses a

controlled list.

- External Links: Can link to Unicode script charts or epigraphic databases.

Geographic/Place Authority File

- Purpose: Standardizes place names for manuscript origins, locations, or cultural contexts to resolve ambiguities.
- Examples:
 - Canonical Name: Varanasi, Madurai.
 - Variant Names: E.g., “Kashi” for Varanasi.
 - Geographic Coordinates: Latitude/longitude for precise mapping.
 - Unique Identifier: E.g., GeoNames ID (e.g., GeoNames:1264542 for Varanasi).
- Relevance to Manuscripts: Indian manuscripts often originate from specific regions (e.g., Kerala for palm-leaf texts). This file supports provenance tracking and relational metadata (e.g., linking manuscripts from the same region). NMM’s “Location” field uses place names.
- External Links: Connects to GeoNames or TGN (Thesaurus of Geographic Names) for global mapping.

Organization/Institution Authority File

- Purpose: Standardizes names of repositories, libraries, or custodians holding manuscripts.
- Examples:
 - Canonical Name: Bhandarkar Oriental Research Institute.
 - Variant Names: E.g., “BORI, Pune.”
 - Unique Identifier: E.g., ISIL (International Standard Identifier for Libraries).
 - Contact Information: Address, repository type (e.g., academic, religious).
- Relevance to Manuscripts: Indian manuscripts are scattered across institutions, temples, and private collections. This file ensures

accurate attribution and supports access control metadata. NMM's "Organization" field tracks custodians.

- External Links: Links to ISIL or WorldCat for repository networking.

Material Authority File

- Purpose: Standardizes terms for manuscript materials to support preservation and description.
- Examples:
 - Canonical Term: Palm leaf, Birch bark, Paper.
 - Variant Terms: E.g., "Talapatra" for palm leaf.
 - Conservation Notes: E.g., "Susceptible to insect damage."
- Relevance to Manuscripts: Indian manuscripts use diverse materials, each requiring specific conservation. This file ensures consistent terminology in NMM's "Material" field, aiding preservation metadata.
- External Links: Can reference conservation standards (e.g., ICOM guidelines).

Relation Authority File

- Purpose: Standardizes terms for relationships between manuscripts to support full-blown relational metadata.
- Examples:
 - Relation Types: IsCommentaryOf, IsCopyOf, IsPartOf, HasTranslation, QuotedIn.
 - Unique Identifier: E.g., TEI's ID or RDF URI.
 - Target Manuscript: Links to another manuscript's identifier (e.g., Manus Id: MG-12345).
- Relevance to Manuscripts: Indian manuscripts often have complex intertextual relationships (e.g., a Sanskrit text with multiple regional commentaries). This file supports TEI's or NMM's "Commentary" field to map these links.
- External Links: Aligns with RDF or CIDOC-CRM for semantic web integration.

Catalogue Authority File

- Purpose: Standardizes references to external or historical catalogues for cross-referencing.
- Examples:
 - Canonical Name: Aufrecht's Catalogus Catalogorum.
 - Variant Names: E.g., "Aufrecht's Catalogue."
 - Unique Identifier: E.g., ISBN or digital catalogue URI.
- Relevance to Manuscripts: Many Indian manuscripts are referenced in historical catalogues. This file, used in NMM's "Catalogue Source" field, supports linking to external resources.
- External Links: Connects to digital archives like HathiTrust or WorldCat.

Implementation in Repositories

- NMM's Approach: NMM's Manus Granthavali software uses authority files for names, titles, subjects, languages, scripts, and organizations, ensuring consistency across its 5.2 million cataloged manuscripts. For example, the "Language" field aligns with ISO 639-2, and "Script" uses a controlled list.
- Standards Integration: Authority files should align with global standards (e.g., VIAF, LCSH, GeoNames) to enable interoperability, as seen in collaborations between NMM and IGNCA.
- Maintenance: Regularly update authority files to incorporate new variants or identifiers, using tools like OpenRefine for data cleaning.

Authority files for personal names, titles, subjects, languages, scripts, places, organizations, materials, relations, and catalogues are essential for a robust manuscript metadata scheme. They ensure consistency, support complex relations (e.g., commentaries), and enable internal/external linking, as exemplified by NMM's Kriti Sampada. By adopting standards like VIAF, ISO 639, and TEI, repositories can create a networked, accessible, and ethically sound system for Indian manuscripts.

6. Bibliographic Relations

Metadata is not limited to describing individual resources through attributes such as title, author, or date. Its true strength emerges when relationships

(relations) between resources are explicitly recorded. Relations enrich metadata by moving it beyond flat description to a network of interconnected knowledge, which is especially critical in complex collections such as manuscripts, archives, and digital libraries. Relations transform metadata from a static catalog into a dynamic knowledge system. They provide context, improve discovery, represent complexity, and ensure interoperability — laying the foundation for scholarly research, digital preservation, and AI-driven applications.

Work-Level Relations

hasWork / isWorkOf

- Links between different expressions of the same intellectual work
- *Example:* Ramayana manuscript from Kerala (MS-RAM-KL-001) hasWork the same intellectual content as Ramayana manuscript from Tamil Nadu (MS-RAM-TN-002)
- *Context:* Different regional versions of the same epic

derivedFrom / hasDerivation

- *Source text to commentary/adaptation relationships*
- *Example:* Adhyatma Ramayana (MS-ADH-001) derivedFrom Valmiki Ramayana (MS-VAL-001)
- *Context:* Spiritual interpretation derived from original epic

translationOf / hasTranslation

- *Original text to translated versions*
- *Example:* Kamba Ramayanam in Tamil (MS-KAM-001) translationOf Valmiki Ramayana in Sanskrit (MS-VAL-001)
- *Context:* Tamil rendering of Sanskrit epic
- *Example:* Bhagavad Gita Telugu version (MS-GIT-TEL-001) translatedBy Bammara Pothana
- *Context:* 15th-century Telugu rendering

commentaryOn / hasCommentary

- *Text to its commentaries/glosses*
- *Example:* Sayana's Rigveda Bhashya (MS-SAY-001) commentaryOn Rigveda Samhita (MS-RIG-001)

- *Context: 14th-century commentary on Vedic hymns*

abridgmentOf / hasAbridgment

- *Complete work to condensed versions*
- *Example: Laghu Bharata (MS-LAG-001) abridgmentOf Mahabharata (MS-MAH-001)*
- *Context: Condensed version of the epic for easier study*

paraphraseOf / hasParaphrase

- *Original to paraphrased versions*
- *Example: Bhagavata Purana prose summary (MS-BHA-PRO-001) paraphraseOf Bhagavata Purana verses (MS-BHA-001)*
- *Context: Prose rendering of verse text*

Manuscript-to-Manuscript Relations

copyOf / hasCopy

- *Manuscript copying relationships*
- *Example: Charaka Samhita copy at BORI (MS-CHA-BOR-001) copyOf original at Kerala Sahitya Akademi (MS-CHA-KSA-001)*
- *Context: 18th-century copy of 16th-century original*

exemplarOf / hasExemplar

- *Master copy relationships*
- *Example: Hatha Yoga Pradipika exemplar (MS-HAT-EX-001) served as model for multiple copies across North India*
- *Context: Master copy used by scribal schools*

fragmentOf / hasFragment

- *Part-to-whole relationships*
- *Example: Single palm leaf from Panchatantra (MS-PAN-FR-001) fragmentOf complete manuscript (MS-PAN-COM-001)*
- *Context: Damaged manuscript with surviving leaves*

boundWith / isBoundWith

- *Part-to-whole relationships*
- *Example: Gita Govinda (MS-GIT-001) boundWith Jayadeva's other*

works (MS-JAY-001) in composite manuscript

- *Context: Multiple texts by same author bound together*

Collection Relations

partOf / hasPart

- *Collection membership*
- *Example: Individual Upanishad manuscript (MS-UPA-ISH-001) partOf larger Upanishad collection (COL-UPA-001)*
- *Context: Systematic collection of related texts*

seriesOf / hasSeries

- *Sequential relationships*
- *Example: Ashvalayana Grihya Sutra Volume 2 (MS-ASH-V2-001) seriesOf multi-volume work*
- *Context: Sequential manuscript volumes*

Intertextual Relations

quotesFrom / isQuotedIn

- *Citation relationships*
- *Example: Sankara's Brahma Sutra Bhashya (MS-SAN-BSB-001) quotesFrom various Upanishads*
- *Context: Philosophical commentary citing source texts*

alludesTo / isAlludedToIn

- *Reference relationships*
- *Example: Kalidasa's Meghaduta (MS-KAL-MEG-001) alludesTo Ramayana episodes*
- *Context: Literary work referencing epic narratives*

respondsTo / hasResponse

- *Reference relationships*
- *Example: Madhva's Dvaita philosophy texts (MS-MAD-001) respondsTo Shankara's Advaita works (MS-SAN-001)*
- *Context: Philosophical debate through texts*

refutesFrom / isRefutedIn

- *Argumentative relationships*
- *Example: Ramanuja's Sri Bhashya (MS-RAM-SRI-001) refutesFrom certain Advaita positions*
- *Context: Sectarian philosophical disputes*
- *Structural Relations*

precededBy / follows

- *Sequential ordering*
- *Example: Aranyakanda manuscript (MS-RAM-ARN-001) precededBy Ayodhyakanda (MS-RAM-AYO-001)*
- *Context: Sequential books of Ramayana*

illustratedBy / illustrates

- *Text-image relationships*
- *Example: Bhagavata Purana text (MS-BHA-TXT-001) illustratedBy miniature paintings (MS-BHA-ILL-001)*
- *Context: Pahari school illustrated manuscripts*

glossedBy / glosses

- *Explanatory relationships*
- *Example: Amarakosha main text (MS-AMA-001) glossedBy interlinear Malayalam explanations*
- *Context: Sanskrit lexicon with vernacular glosses*

AUTHORSHIP AND ATTRIBUTION RELATIONS

authoredBy / isAuthorOf

- *Primary authorship*
- *Example: Gita Govinda (MS-GIT-001) authoredBy Jayadeva (12th century)*
- *Context: Established authorship attribution*

attributedTo / isAttributedWork

- *Traditional attribution*
- *Example: Viveka Chudamani (MS-VIV-001) attributedTo Adi Shankara*

- *Context: Traditional attribution, scholarly debate ongoing*

compiledBy / isCompilationOf

- *Editorial compilation*
- *Example: Panchatantra collection (MS-PAN-001) compiledBy Vishnu Sharma*
- *Context: Story collection compiled from oral traditions*

translatedBy / isTranslationOf

- *Translation work*
- *Example: Bhagavad Gita Telugu version (MS-GIT-TEL-001) translatedBy Bammara Pothana*
- *Context: 15th-century Telugu rendering*

Scribal Relations

scribedBy / isScribeWork

- *Copyist relationships*
- *Example: Hastāmalaka Stotra manuscript (MS-HAS-001) scribedBy Narayana Bhatta (dated 1654 CE)*
- *Context: Identified scribe with colophon information*

copiedBy / isCopyOf

- *Copying relationships*
- *Example: Yoga Vasishtha copy (MS-YOG-KAS-001) copiedBy Kashmir scribes at Sharada Peetha*
- *Context: Institutional copying tradition*

correctedBy / isCorrectionOf

- *Correction work*
- *Example: Sushruta Samhita manuscript (MS-SUS-001) correctedBy later scholarly hand*
- *Context: Editorial corrections in different ink*

Patronage Relations

patronizedBy / isPatronageOf

- *Patron relationships*

- *Example: Ragamala manuscript (MS-RAG-001) patronizedBy Maharaja Sawai Jai Singh II of Jaipur*
- *Context: Royal patronage of musical texts*

commissionedBy / isCommissionOf

- *Commission relationships*
- *Example: Hamzanama illustration series (MS-HAM-001) commissionedBy Akbar's court*
- *Context: Mughal imperial commissioning*

dedicatedTo / isDedicationOf

- *Dedication relationships*
- *Example: Kavyaprakasha commentary (MS-KAV-001) dedicatedTo Vijayanagara king Krishnadevaraya*
- *Context: Literary work dedicated to royal patron*

Historical and Provenance Relations

Temporal Relations

contemporaryWith / isContemporaryOf

- *Same time period*
- *Example: Tulsidas's Ramcharitmanas (MS-TUL-001) contemporaryWith Akbar's reign manuscripts*
- *Context: 16th-century contemporaneous works*

dynastyOf / hasDynasty

- *Dynastic period*
- *Example: Chola period Tamil inscriptions (MS-CHO-001) dynastyOf Chola dynasty (9th-13th centuries)*
- *Context: Historical period classification*

Ownership Relations

ownedBy / isOwnerOf

- *Ownership relationships*
- *Example: Natya Shastra manuscript (MS-NAT-001) ownedBy Thanjavur Saraswathi Mahal Library*

- *Context: Current institutional ownership*

inheritedBy / isInheritanceOf

- *Inheritance relationships*
- *Example: Family Ayurveda collection (MS-AYU-FAM-001) inheritedBy traditional Vaidya family in Kerala*
- *Context: Hereditary manuscript transmission*

donatedTo / isDonationOf

- *Gift relationships*
- *Example: Jain Kalpasutra manuscript (MS-JAI-KAL-001) donatedTo L.D. Institute by Gujarati merchant family*
- *Context: Community donation to research institution*

Geographical Relations

Spatial Relations

originatesFrom / isOriginOf

- *Place of origin*
- *Example: Sharada script Kashmiri manuscripts (MS-SHA-001) originatesFrom Kashmir valley*
- *Context: Regional script and content origin*

producedIn / isProductionSiteOf

- *Production location*
- *Example: Tanjore style illustrated Ramayana (MS-TAN-RAM-001) producedIn Thanjavur court workshops*
- *Context: Specific production location*

foundIn / isFindSiteOf

- *Discovery location*
- *Example: Ancient Buddhist manuscript fragments (MS-BUD-FR-001) foundIn Gilgit archaeological site*
- *Context: Archaeological discovery location*

Cultural Geographic Relations

regionOf / hasRegion

- *Regional classification*
- *Example: Manippravalam literature manuscripts (MS-MAN-001) regionOf Kerala region*
- *Context: Regional literary tradition*

cultureOf / hasCulture

- *Cultural area*
- *Example: Jain Agama manuscripts (MS-JAI-AGA-001) cultureOf Gujarati Jain community*
- *Context: Community-specific religious texts*

Language Relations

writtenIn / isLanguageOf

- *Primary language*
- *Example: Thirukkural manuscript (MS-THI-001) writtenIn Tamil language*
- *Context: Primary language identification*

bilingualWith / isBilingualOf

- *Multi-language texts*
- *Example: Sanskrit-Tamil bilingual Shaiva Agama (MS-SHA-AGA-001) with parallel texts*
- *Context: Bilingual religious manuscripts*

dialectOf / hasDialect

- *Language variants*
- *Example: Maithili manuscript (MS-MAI-001) dialectOf Hindi language family*
- *Context: Regional language variant*

Script Relations

scriptOf / hasScript

- *Writing system*
- *Example: Grantha script Vaishnavite manuscript (MS-GRA-VAI-001) used for Sanskrit in South India*

- *Context: Regional script usage*

paleographyOf / hasPaleography

- *Script development*
- *Example: 12th-century Devanagari script (MS-DEV-12C-001) showing transitional forms*
- *Context: Script evolution documentation*

Thematic and Subject Relations

Subject Relations

subjectOf / hasSubject

- *Topic relationships*
- *Example: Charaka Samhita (MS-CHA-001) subjectOf Ayurvedic medicine*
- *Context: Medical treatise subject classification*

disciplineOf / hasDiscipline

- *Academic field*
- *Example: Siddhanta Shiromani (MS-SID-001) disciplineOf mathematical astronomy (Jyotisha)*
- *Context: Academic discipline classification*

traditionOf / hasTradition

- *Literary tradition membership*
- *Example: Tantric manuscript (MS-TAN-001) traditionOf Shakta tradition*
- *Context: Religious tradition membership*

Conceptual Relations

philosophyOf / hasPhilosophy

- *Philosophical system*
- *Example: Yoga Sutras commentary (MS-YOG-SUT-001) philosophyOf Samkhya-Yoga darshana*
- *Context: Philosophical system classification*

theologyOf / hasTheology

- *Religious doctrine*
- *Example: Vaishnava theological manuscript (MS-VAI-THE-001) theologyOf devotional theology*
- *Context: Religious doctrine classification*

Material Relations

madeFrom / isMaterialOf

- *Substrate materials*
- *Example: Kerala palm leaf manuscript (MS-PAL-KER-001) madeFrom Palmyra palm leaves*
- *Context: Traditional writing substrate*

inkOf / hasInk

- *Writing materials*
- *Example: Rajasthani miniature manuscript (MS-RAJ-MIN-001) inkOf mineral pigments and gold leaf*
- *Context: Traditional ink and paint materials*

bindingOf / hasBinding

- *Binding materials*
- *Example: Mughal manuscript (MS-MUG-001) bindingOf leather with gold tooling*
- *Context: Traditional Islamic binding style*

Technical Relations

techniqueOf / hasTechnique

- *Production methods*
- *Example: Jain manuscript illumination (MS-JAI-ILL-001) techniqueOf Western Indian painting style*
- *Context: Regional artistic technique*

processOf / hasProcess

- *Manufacturing processes*
- *Example: Paper manuscript (MS-PAP-001) processOf traditional Indian paper-making from bamboo*

- *Context: Indigenous manufacturing process*

Preservation and Conservation Relations

Condition Relations

damagedBy / isDamageSource

- *Damage causation*
- *Example: Monsoon-affected manuscript (MS-MON-001) damagedBy Kerala tropical climate*
- *Context: Environmental damage factors*

treatedBy / isTreatmentOf

- *Conservation work*
- *Example: Restored Saraswathi Mahal manuscript (MS-SAR-001) treatedBy National Research Laboratory for Conservation*
- *Context: Professional conservation work*

digitizedBy / isDigitizationOf

- *Digital conversion*
- *Example: Online Mahabharata manuscript (DIG-MAH-001) digitizedBy Bhandarkar Oriental Research Institute*
- *Context: Digital preservation project*

Scholarly and Research Relations

studiedBy / isStudyOf

- *Research relationships*
- *Example: Panini's Ashtadhyayi manuscript (MS-PAN-ASH-001) studiedBy multiple Sanskrit grammatical research projects*
- *Context: Academic research focus*

catalogedBy / isCatalogOf

- *Cataloging work*
- *Example: Descriptive catalogue entry (CAT-001) catalogedBy Dr. V. Raghavan for Adyar Library manuscripts*
- *Context: Scholarly cataloging work*

publishedBy / isPublicationOf

- *Publication relationships*
- *Example: Critical edition (PUB-CRI-001) publishedBy Oriental Institute, Baroda based on multiple manuscripts*
- *Context: Scholarly publication based on manuscript sources*

Comparative Relations

comparedWith / isComparisonOf

- *Comparative studies*
- *Example: Northern Ramayana recension (MS-RAM-NOR-001) comparedWith Southern recension (MS-RAM-SOU-001)*
- *Context: Textual criticism and comparison*

Ceremonial and Ritual Relations

Functional Relations

usedFor / isFunctionOf

- *Functional purposes*
- *Example: Vedic ritual manual (MS-VED-RIT-001) usedFor Agnihotra ceremonies*
- *Context: Ritual performance texts*

performedIn / isPerformanceVenue

- *Performance contexts*
- *Example: Kathakali performance text (MS-KAT-001) performedIn Kerala temple festivals*
- *Context: Performance tradition venues*

liturgyOf / hasLiturgy

- *Liturgical usage*
- *Example: Tamil Shaiva hymns manuscript (MS-SHA-HYM-001) liturgyOf temple worship services*
- *Context: Religious liturgical usage*

Sacred Relations

consecratedFor / isConsecrationOf

- *Sacred dedication*
- *Example: Temple installation manual (MS-TEM-INS-001) consecratedFor deity consecration rituals*
- *Context: Sacred ceremonial usage*

blessedBy / isBlessingOf

- *Religious blessing*
- *Example: Guru parampara manuscript (MS-GUR-PAR-001) blessedBy lineage head before transmission*
- *Context: Traditional blessing of sacred texts*

7. Probable Role of Artificial Intelligence in Metadata

Generating Metadata from Unstructured Data

AI can leverage natural language processing (NLP), optical character recognition (OCR), and machine learning (ML) to extract metadata elements (e.g., title, author, language, relations like hasWork or commentaryOn) from unstructured manuscript text. The process involves:

- **Digitization and OCR:** IndicOCR (AI4Bharat) and Google Cloud Vision AI with Indic Plugins may Supports Indic languages and scripts
- **Language and Script Identification:** IndicLID (AI4Bharat) and ILID (Native Script LID) may handle cases where multiple languages share scripts (e.g., Hindi/Marathi/Sanskrit in Devanagari).
- **Entity and Relation Extraction:** NLP techniques like named entity recognition (NER) and relation extraction identify metadata elements (e.g., title, author, place) and relations (e.g., hasWork, commentaryOn) from unstructured text, such as colophons or body text. Tools like spaCy fine-tuned for Indic languages.
- **External Resource Enrichment:** AI cross-references extracted metadata with external sources (e.g., Wikidata, VIAF) to validate and enrich fields like author names or relations (e.g., hasWork). Tools: OpenRefine: Reconciles metadata with authority files (e.g., VIAF for “Jayadeva,” GeoNames for “Kerala”). Wikidata SPARQL: Queries for relations like hasWork or authoredBy.
- **Validation and Human Oversight:** AI flags uncertain extractions

(e.g., ambiguous relations like refutesFrom) for human review, using confidence scores from NLP models. Oxygen XML Editor: Validates TEI schemas. AI flags “Ramanuja’s Sri Bhashya refutes Advaita” for review due to philosophical nuance.

AI based Text Summarisation

The manual summarization is labour-intensive due to the multilingual and multiscript nature of the content. AI generates extractive (selecting key sentences) or abstractive (paraphrasing in new sentences) summaries. Multilingual models handle Indic languages, producing outputs in the original language, English, or other regional languages. Example Tools:

- AI4Bharat’s IndicGenBench and Models (e.g., IndicBERT, IndicBART): Benchmarks and generates summaries for 29 Indic languages (high-resource like Hindi/Tamil, low-resource like Manipuri/Santali) across tasks like cross-lingual summarization. Supports scripts like Devanagari, Bengali, Tamil, and Meitei. For instance, it can summarize a Sanskrit manuscript in Hindi or English.
- mBART or Indic-Specific Transformers (Hugging Face): Fine-tuned for abstractive summarization in Indic languages (e.g., Hindi, Tamil, Sanskrit). Tools like Extractive Text Summarization pipelines handle English/Hindi/Tamil directly, using frequency-based scoring for sentences.

Challenges and Limitations

- Script/OCR Accuracy: Ancient scripts like Sharada or Grantha may have lower OCR accuracy; tools like IndicOCR are improving but require fine-tuning.
- Low-Resource Languages: Summarization for languages like Santali or Maithili is less accurate due to limited training data, but benchmarks like IndicGenBench help evaluate and improve.
- Contextual Nuance: Philosophical manuscripts (e.g., Vedanta commentaries) may lose subtlety in AI summaries; human oversight is needed.
- Ethical Considerations: AI must respect cultural sensitivities (e.g., sacred texts); NMM’s focus on preservation ensures summaries don’t alter originals.

In summary, AI tools not only can but are already being developed and applied (e.g., via AI4Bharat, Vikas AI) to generate abstracts for NMM manuscripts, supporting India's linguistic diversity.

Querying the Database using Natural Language

AI tools can effectively answer natural language queries against a database of metadata records, provided the database is structured and accessible. AI tools can interpret queries, map them to structured metadata fields and retrieve or reason over the data to provide accurate responses.

- **Query Interpretation with NLP:** AI parses natural language queries using NLP techniques, identifying key entities (e.g., titles, authors, relations like “commentaryOn”), intents (e.g., search, compare), and constraints (e.g., “Sanskrit manuscripts”). Tokenization, named entity recognition (NER), and dependency parsing extract relevant components.
- **Example:** For the query “Which manuscripts are commentaries on the Bhagavad Gita?”, the AI identifies “commentaries” (relation: commentaryOn), “Bhagavad Gita” (target text), and “manuscripts” (entity type).
- **Response Generation:** The AI synthesizes results into a natural language response, summarizing metadata or providing detailed records. It can include context (e.g., historical significance) or visualizations (e.g., network graphs for relations like hasSeries).
- **Example:** “Manuscripts commenting on the Bhagavad Gita include Sankara Bhashya (MS-SAN-001) and Ramanuja Bhashya (MS-RAM-001), both in Sanskrit, held at BORI.”
- **Tools and Technologies:** spaCy, NLTK, or transformer-based models (e.g., BERT) for query parsing.

Types of Queries AI Can Answer

- Query: “Which manuscripts are different regional versions of the Ramayana?”
- Relation: hasWork / isWorkOf
- Query: “Which manuscripts are 18th-century copies of older originals?”
- Relation: copyOf / hasCopy

- Query: “What are the surviving fragments of the Panchatantra?”
- Relation: fragmentOf / hasFragment
- Query: “Which manuscripts quote from the Upanishads?”
- Relation: quotesFrom / isQuotedIn
- Query: “Which manuscripts are made from palm leaves?”
- Relation: madeFrom / isMaterialOf

AI tools like me can answer a wide range of natural language queries against manuscript metadata databases, leveraging relations like hasWork, commentaryOn, or digitizedBy. By parsing queries, mapping to structured metadata and retrieving results, AI supports discovery, intertextual analysis, and provenance tracking for Indian manuscripts. Challenges like incomplete metadata or multilingual complexity can be mitigated with enriched schemas and NLP advancements.

How AI Tools Assist in Identifying Relations

AI tools streamline data entry by automating relation identification, validation, and encoding, reducing manual effort while improving accuracy and scalability. Here’s the process:

- Text Analysis and Entity Recognition:
 - Process: AI uses NLP techniques like named entity recognition (NER), relation extraction, and text classification to identify entities (e.g., titles, authors, places) and relations (e.g., commentaryOn, derivedFrom) from manuscript texts, colophons, or existing metadata. Pre-trained models (e.g., BERT, spaCy) or fine-tuned models for manuscript-specific vocabularies (e.g., Sanskrit terms) are used.
 - Example: From a colophon stating “Sankara’s commentary on Bhagavad Gita,” AI extracts “Sankara Bhashya” (entity: manuscript), “Bhagavad Gita” (entity: base text), and “commentary” (relation: commentaryOn).
- Relation Extraction and Classification:
 - Process: AI models, such as transformer-based relation extractors or rule-based systems, classify relationships between entities (e.g., hasWork, copyOf). Supervised learning with labeled datasets (e.g., annotated TEI records) or unsupervised methods (e.g., clustering

similar texts) identify relations like quotesFrom or patronizedBy.

- Example: AI detects “Kamba Ramayanam is a Tamil rendering of Valmiki Ramayana” and classifies it as translationOf, linking MS-KAM-001 to MS-VAL-001.
- Metadata Encoding:
 - Process: AI generates structured metadata (e.g., TEI XML `<relatedItem>`, CIDOC-CRM triples) based on extracted relations, inserting them into the repository’s database (e.g., eXist-db, RDF triplestore). Templates ensure compliance with standards like TEI P5 or Dublin Core.
 - Example: AI encodes `<relatedItem type=”commentary” target=”#MS-GIT-001”/>` for Sankara Bhashya in a TEI record.
- External Resource Integration:
 - Process: AI crawls and analyzes Internet resources (e.g., digital catalogues, Wikidata, scholarly articles) to identify relations, using APIs or web scraping. It matches entities with authority files (e.g., VIAF for authors, GeoNames for places) to link internal records to external data.
 - Example: AI finds a British Library catalogue entry linking a Persian Upanishads translation to a Sanskrit original, encoding it as hasTranslation.
- Validation and Human Oversight:
 - Process: AI suggests relations with confidence scores, flagging uncertain cases for human review. Active learning refines models with curator feedback, improving accuracy for complex relations like refutesFrom or isConsecrationOf.
 - Example: AI suggests “Ramanuja’s Sri Bhashya refutes Advaita” but flags for review due to philosophical nuance, ensuring curator validation.

AI Capabilities for External Resource Integration

AI can connect repository metadata to external Internet resources to enrich relations, leveraging digital catalogs, scholarly databases, and linked open data. Methods include:

- Web Crawling and Scraping:
 - Process: AI crawls digital libraries (e.g., British Library, IGNCA) or catalogs (e.g., Aufrecht's Catalogus Catalogorum) to find related manuscripts, using APIs or scraping tools like BeautifulSoup.
 - Example: AI finds a British Library record linking a Persian Upanishads manuscript (MS-UPA-PER-001) as hasTranslation to a Sanskrit original, encoding it in the repository.
- Linked Open Data Integration:
 - Process: AI queries RDF datasets (e.g., Wikidata, Library of Congress) using SPARQL to match entities and relations (e.g., hasWork, patronizedBy).
 - Example SPARQL Query:

```

```sparql
SELECT ?manuscript ?relation ?target
WHERE {
 ?manuscript wdt:P629 ?work . # hasWork relation
 ?work wdt:P1476 "Ramayana" .
 ?manuscript ?relation ?target .
}
```

```

Result: Links Kerala Ramayana (MS-RAM-KL-001) to Tamil Nadu Ramayana (MS-RAM-TN-002) via Wikidata's work entity.

- Authority File Matching:
 - Process: AI matches authors, places, or works to VIAF, GeoNames, or NMM's Manus Id for consistency.
 - Example: AI links "Jayadeva" in Gita Govinda (MS-GIT-001) to VIAF ID, confirming authoredBy relation.

AI tools excel at identifying and encoding relations within a manuscript repository and from external Internet resources, automating data entry for relations like hasWork, commentaryOn, or originatesFrom. By using NLP, ML, and linked data, AI extracts relations from texts, metadata, and online sources, encoding them in standards like TEI or CIDOC-CRM. Challenges

like multilingual texts or incomplete data can be addressed with fine-tuned models and human oversight.

Challenges and Solutions

- **Complex Relations:**
 - Issue: Relations like `refutesFrom` or `isConsecrationOf` require nuanced understanding.
 - Solution: Use fine-tuned NLP models with domain-specific training (e.g., Indic philosophical terms) and curator validation.
- **Metadata Inconsistency:**
 - Issue: Incomplete or inconsistent relation encoding (e.g., missing `<relatedItem>`).
 - Solution: AI infers relations from text or cross-references external sources (e.g., Wikidata) to fill gaps.
- **Multilingual Data:**
 - Issue: Indian manuscripts use Sanskrit, Tamil, etc., complicating parsing.
 - Solution: Employ multilingual NLP models (e.g., mBERT) and Unicode for scripts.
- **Overloading the Abstract:**
 - Issue: Including all relations (e.g., 40+ from the list) can make abstracts unwieldy.
 - Solution: Prioritize key relations (e.g., `hasWork`, `authoredBy`) or categorize into sections (textual, provenance, material).

Tools and Technologies

- **Metadata Parsing:**
 - XML Parsers: `lxml`, `ElementTree` for TEI XML.
 - RDF Querying: Apache Jena, SPARQL for CIDOC-CRM.
 - Databases: eXist-db, BaseX for TEI; SQL for Dublin Core.

- NLP for Summarization:
 - Transformers: Hugging Face's T5 or BART for abstractive summarization.
 - spaCy: For entity and relation extraction.
- External Enrichment:
 - APIs: Wikidata, VIAF for author/place verification.
 - Web Crawling: BeautifulSoup for digital catalogs.
- Validation:
 - Oxygen XML Editor: Validate TEI schemas.
 - Active Learning: Refine summaries with curator feedback.

Ensuring Error-Free metadata

While no single tool guarantees completely error-free metadata, a combination of metadata cleaning tools can address issues like inconsistencies, duplicates, missing values, and incorrect relations. Below, I provide a list of open source tools designed for metadata cleaning, validation, and standardization, focusing on their applicability to structured metadata

Criteria for Metadata Cleaning Tools

To ensure error-free metadata, tools should:

- Identify Errors: Detect inconsistencies, duplicates, missing values, or incorrect relations (e.g., mismatched commentaryOn targets).
- Validate Against Standards: Check compliance with schemas like TEI P5, Dublin Core, or CIDOC-CRM.
- Standardize Data: Normalize formats (e.g., dates, names) using authority files (e.g., VIAF, GeoNames).
- Support Batch Processing: Handle large datasets, common in manuscript repositories.
- Preserve Data Integrity: Ensure original content (e.g., manuscript text) is unaffected.
- Handle Multilingual Data: Support Indic languages (e.g., Sanskrit, Tamil)

and scripts (e.g., Devanagari).

A List of Open Source Metadata Cleaning Tools

- OpenRefine: A powerful tool for cleaning and transforming messy metadata, widely used for data reconciliation and standardization.
 - Features:
 - Data exploration via facets and charts to identify inconsistencies (e.g., variant spellings of “Shankara”).
 - Transformation using GREL (General Refine Expression Language) for normalizing fields (e.g., dates to ISO 8601).
 - Reconciliation with external sources (e.g., Wikidata, VIAF) to standardize author names or manuscript titles.
 - Clustering to merge duplicates (e.g., “Bhagavad Gita” vs. “Bhagvat Gita”).
 - Supports CSV, JSON, XML, and TEI XML (via plugins).
 - Applicability: Ideal for cleaning NMM metadata fields (e.g., title, author) and relations (e.g., hasWork) by reconciling with authority files. Supports batch processing for large repositories.
 - Limitations: Limited support for large datasets (>1M records); advanced TEI relation cleaning requires scripting.
 - Use Case: Normalize inconsistent script names (e.g., “Devanagari” vs. “Devanāgarī”) and validate ``<relatedItem>`` targets.
- ExifTool: A versatile CLI tool for reading, writing, and editing metadata in a wide range of file formats.
- mat2 (Metadata Anonymization Toolkit) : A Python-based CLI tool for removing metadata from images, videos, documents, and torrents.

Conclusion

The preservation and scholarly engagement with India’s vast manuscript heritage—estimated at over ten million manuscripts—demands a robust, intelligent, and future-ready metadata framework. This paper attempts to present a fairly comprehensive vision for such a system through the proposed metadata scheme for Gyan Bharatam, emphasizing structured description,

rich relational modelling, and the integration of Artificial Intelligence to meet the unique challenges posed by Indian manuscripts. The salient contribution lies not merely in cataloguing manuscripts, but in transforming them into an interconnected, semantically rich, and intellectually accessible knowledge network.

At the heart of this framework is the recognition that metadata is far more than administrative overhead—it is the key to discovery, preservation, interpretation, and cultural continuity. Indian manuscripts, inscribed in diverse scripts such as Devanagari, Grantha, and Sharada, composed in multiple languages including Sanskrit, Tamil, and Persian, and produced on materials ranging from palm leaves to handmade paper, require metadata that captures their physical, linguistic, historical, and cultural complexity. The evaluation of the National Mission for Manuscripts (NMM) metadata schema reveals commendable strengths in descriptive and administrative documentation, particularly in its adaptation of Dublin Core to Indian contexts and its focus on Indic scripts via Unicode. However, gaps remain in relational depth, digital preservation standards, and integration with global authority systems—gaps that must be addressed to enable advanced research and interoperability.

A defining feature of this proposal is the extensive taxonomy of bibliographic and contextual relations—over forty types spanning work-level connections (e.g., `hasWork`, `translationOf`), manuscript-to-manuscript links (`copyOf`, `fragmentOf`), intertextual references (`quotesFrom`, `respondsTo`), and cultural-religious functions (`usedFor`, `consecratedFor`). These relational elements are not ancillary; they are essential for reconstructing intellectual lineages, tracing textual evolution, and understanding the socio-cultural ecosystems in which manuscripts were created and used. By adopting standards such as TEI and CIDOC-CRM, and linking entities through persistent identifiers (VIAF, DOI, GeoNames), this metadata model moves beyond flat records toward a dynamic, queryable knowledge graph.

Equally transformative is the role of Artificial Intelligence in enhancing metadata creation and access. AI tools—powered by multilingual NLP models like those from AI4Bharat—can automate the extraction of metadata from unstructured text, identify complex relations in colophons or commentaries, generate multilingual summaries, and reconcile entities with global authority files. Natural language querying allows scholars to ask, “Which 18th-century manuscripts are commentaries on the Bhagavad Gita?” and receive precise, context-aware responses. Furthermore, AI facilitates external enrichment by linking repository data to resources like Wikidata, WorldCat, and digital archives, creating a globally connected ecosystem of knowledge.

To ensure accuracy and cultural sensitivity, human oversight remains indispensable. AI must operate in tandem with domain experts, especially

when dealing with philosophical nuances, contested attributions, or sacred texts requiring ethical handling. Tools like OpenRefine, Apache Jena, and Oxygen XML Editor support data cleaning, validation, and standardization, ensuring that metadata is both machine-actionable and scholar-trustworthy.

The preservation and scholarly engagement with India's vast manuscript heritage—estimated at over ten million manuscripts—demands a robust, intelligent, and future-ready metadata framework. This paper has presented a comprehensive vision for such a system through the proposed metadata scheme for Gyan Bharatam, emphasizing structured description, rich relational modeling, and the integration of Artificial Intelligence to meet the unique challenges posed by Indian manuscripts. The salient contribution lies not merely in cataloguing manuscripts, but in transforming them into an interconnected, semantically rich, and intellectually accessible knowledge network.

Caveat: While Artificial Intelligence is making significant contributions in many domains, it is still far from fully mature. Developing a system capable of handling complex relationships requires advanced machine learning tools to ensure increasing robustness over time. To be comprehensive, this paper enumerates about 40 different types of relations, some of which are not inherently bidirectional. Furthermore, it is strongly recommended that an expert committee—comprising specialists from diverse fields such as Indology, Linguistics, and Vedic, Buddhist, and Jain studies—be constituted to guide this process. Most importantly, establishing a robust metadata schema prior to populating Gyan Bharatam is essential, as revising or re-entering metadata later is far more time-consuming than entering it correctly at the outset.

In conclusion, the proposed metadata framework for Gyan Bharatam represents a paradigm shift—from static archives to intelligent, relational, and semantically enriched knowledge systems. It bridges tradition and technology, philology and data science, local custodianship and global access. By embracing relational depth, standardized authority control, and AI-driven augmentation, India can not only preserve its manuscript legacy but also unlock its full intellectual potential for generations to come.

References:

1. National Mission for Manuscripts (NMM). (2023). Guidelines for Digitization of Manuscripts. New Delhi: Ministry of Culture, Government of India.
2. Dublin Core Metadata Initiative (DCMI). (2012). Dublin Core Metadata

- Element Set, Version 1.1. <https://www.dublincore.org/specifications/dublin-core/dces/>
3. Text Encoding Initiative (TEI). (2023). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Edited by Lou Burnard and Syd Bauman. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
 4. CIDOC Conceptual Reference Model (CRM). (2022). Official CRM Standard. International Council of Museums (ICOM), Documentation Committee. <https://www.cidoc-crm.org/>
 5. AI4Bharat. (2023). IndicOCR, IndicLID, and IndicGenBench: Open-Source Tools for Indian Languages and Scripts. Indian Institute of Technology Madras. <https://ai4bharat.org>
 6. Library of Congress. (2023). MARC Standards. <https://www.loc.gov/standards/marc/>
 7. ISO 639-2.(1998). Codes for the Representation of Names of Languages—Part 2: Alpha-3 Code. International Organization for Standardization.
 8. GeoNames. (2023). GeoNames Geographic Database. <http://www.geonames.org>
 9. Virtual International Authority File (VIAF). (2023). <https://viaf.org>
 10. OpenRefine. (2023). Open Source Tool for Data Cleaning and Transformation. <https://openrefine.org>
 11. Aufrecht, W. (1891–1896). *Catalogus Catalogorum: An Alphabetical Register of Sanskrit Works and Authors* (Vols. 1–3). Leipzig: Breitkopf & Härtel.
 12. UNESCO. (2005). *Multilingual Thesaurus on Cultural Heritage*. UNESCO Publishing.
 13. WorldCat. (2023). Global Library Catalog. OCLC. <https://www.worldcat.org>
 14. HathiTrust Digital Library. (2023). <https://www.hathitrust.org>
 15. Wikidata. (2023). Free and Open Knowledge Graph. <https://www.wikidata.org>
 16. Apache Jena. (2023). Open Source Semantic Web Framework for Java. <https://jena.apache.org>

17. spaCy. (2023). Open-source library for advanced Natural Language Processing in Python. <https://spacy.io>
18. Hugging Face. (2023). Transformers: State-of-the-art Natural Language Processing. <https://huggingface.co>
19. Oxygen XML Editor. (2023). XML Authoring and Development Tool. Syncro Soft. <https://www.oxygenxml.com>
20. National Research Laboratory for Conservation of Cultural Property (NRLC). (2023). Lucknow, India. <http://nrlc.gov.in>
21. Indira Gandhi National Centre for the Arts (IGNCA). (2023). New Delhi, India. <https://ignca.gov.in>

Epilogue

Over the past couple of weeks, I had the privilege of listening to Sri Maharaj ji and Dr. Sanjay Singhal, and I was deeply impressed by the remarkable work being undertaken at the Koba Centre. The Centre's extensive experience in manuscript studies and the strong conceptual framework behind the Sangraha site provide valuable insights that could be effectively leveraged in shaping Gyan Bharatam. The metadata model outlined in this paper can serve as a complementary foundation in building a unified structure for realizing the vision of Gyan Bharatam. Below is a brief overview of the Koba model which should convince anyone having keen interest in the success of Gyan Bhartam.

Koba Manuscript Metadata Search Portal

The Koba Manuscript Metadata Search Portal, developed by Shri Mahavir Jain Aradhana Kendra, Acharya Shri Kailasagarsuri Gyanmandir (Koba), is a pioneering digital initiative that bridges centuries of Indian heritage with cutting-edge technology. Built on a robust metadata framework comprising over 100 searchable fields, it empowers scholars, devotees, and researchers to explore manuscripts and texts with exceptional precision.

At its core, the portal interlinks a wide range of authorities—handwritten manuscripts (Hastaprat/Pandulipi), scholarly works (Kruti), records of eminent scholars (Vidvans), printed publications, magazines, publishers, series, books, cities, and repositories (Bhandars). This interconnected structure creates a dynamic web of knowledge, enabling seamless navigation across related works, persons, and institutions.

Authorities Catalogued in the Koba Database

1. Hastaprat (Manuscripts – Handwritten data)

- Hastaprat Petank
- Pushpika Shlok
- Year Information (Month–Paksha–Tithi,etc.)

2. Kruti (Texts – Works by scholars)

- Text Chapters (Subtitles of texts)
- Adivakya (First sentence of the text)
- Antimvakya (Last sentence of the text)

3. Vidvan (Scholars / Personal Information)

- Gachha details

4. Prakashan (Printed publications)

- Prakashan Petank

5. Magazine

- Individual issues (Magazine Ank)
- Magazine Ank Petank

6. Publisher (Entities responsible for publishing texts)

7. City (Geographical location)

8. Series

9. Book (Physical format of a Prakashan)

10. Bhandar (Libraries and repositories)

Features of the Portal

- Comprehensive Metadata Access: Full bibliographic details of manuscripts, books, and magazines, including titles, page numbers, and chapter details, ensuring accurate citation and research depth.
- Unique URLs: Each entry is assigned a permanent link, allowing precise

citation, sharing, and cross-referencing.

- Multilingual & Fuzzy Search: Supports Gujarati, Hindi, and English for inclusive access.
- Entity-Specific Filters: Enables refined exploration across manuscripts, texts, and authorities.
- PDF Availability Filter: Instantly shows which works are accessible online.
- “Show Blank Fields” Option: Helps users identify records with incomplete metadata for further study or improvement.
- Mobile-Friendly Interface & Dedicated App: Provides global access to one of the world’s largest computerized Indian manuscript collections.

A Proposed Catalogue for Gyan Bharatam Mission

Sanjaya Singhal, Managing Trustee, Dharohar

P. M. Gupta, University Librarian, Central Sanskrit University, New Delhi

Parveen Babbar, Deputy Librarian, JNU

1. Digital catalogue only:

- 1.1. GBM expects to cover **1 crore manuscripts**. The sheer volume of manuscripts makes it imperative that the catalogue is digital.
- 1.2. A digital catalogue has many advantages, the main one being the speed of searching it.

2. Scholars are main users: The primary purpose of the GBM catalogue is to serve Sanskrit scholars worldwide. The catalogue should therefore serve the following two needs of these users:

- 2.1. Allow scholars to find manuscripts of their research interest quickly.
- 2.2. Having identified the required manuscript, research scholars should be able to download digital images of the manuscript from a secure repository of digital images.

3. Participating libraries are important users too: A secondary purpose of the catalogue is to serve participating manuscript collectors and libraries. For this user group the catalogue needs to provide the following:

- 3.1. A complete and accurate inventory of their collection.
- 3.2. Catalogue data of their manuscripts to be included in their own general library catalogue.
- 3.3. Free and unfettered access to digital images of manuscripts of their own collection.
- 3.4. In case a library desires to publish a printed catalogue of their collection, the required metadata in a format suitable of printing.

4. What scholars expect from a manuscript catalogue: Work was done in the recent past with Sanskrit research scholars to identify their user needs. It is likely that research scholars of regional languages will have the same user requirements. The user requirements were expressed in two parts:

- 4.1. The list of metadata needed by Sanskrit scholars to identify

manuscripts their interest. This list is attached at Annex-1.

- 4.2. The methods by which they expect to search the catalogue and view the search results. Since this aspect is not relevant for this stage of the GBM programme, details are not included here.

5. What defines a good catalogue

- 5.1. From the scholars' perspective, the most important requirement is that it should be easy to search manuscripts of their research interest. A search could commence in different ways, e.g. using a title, an author's name, a commentator's name, a subject, a keyword etc. The scholar expects to see results that enable a quick identification of the manuscripts of interest.

- 5.2. Given the expected volume of manuscripts that will be available through GBM, this task becomes quite complex. One title or one author name may reveal hundreds of manuscripts. Therefore, one needs to ensure that the catalogue and the search methods provide precise and complete results.

- 5.2.1. A precise search Is one that gives search results relevant only to the search terms applied i.e. no irrelevant information is provided by a precise search.

- 5.2.2. A complete search is one where all results relevant to the search terms are provided i.e. no relevant manuscript is missed out.

6. Quality of metadata

- 6.1. The quality of search therefore is a direct result of the consistency with which the metadata extracted for each manuscript and the manner in which it is recorded in the bibliographic record of that manuscript.
- 6.2. As a first step achieving good quality requires detailed guidelines for the extraction of metadata. Manuscripts present a very high level of variability and the guidelines need to ensure that they cater to this variability found in the manuscripts.
- 6.3. For an effective computer search, the data needs to be recorded in a consistent manner. Thus, for example, if a book has is known by more than one title, its commonly known title should be recorded as the main title and every cataloguer must always use that same title.

- 6.4. Consistency requires an effective vocabulary / list of authorities to be available for key metadata fields such as title, author, commentator, subject, place of writing etc. Most other metadata fields lend themselves to a choice from a set of acceptable list of values. Choosing the right entry from this list of values will ensure that for most other fields data is entered consistently.
- 6.5. A list of authorities and the data required for each entry is given at Annex-3.
- 6.6. It is worth recording here that a task of the scale of GBM requires a process of catalogue creation and metadata extraction such that quality and consistency is built into the process itself. No amount of supervision or inspection can guarantee quality and consistency at such a large scale.

7. Romanisation

- 7.1. For international scholars providing metadata in a suitable Romanized script is necessary. Romanization may also be necessary for Indian scholars who may or may not be familiar with regional languages.
- 7.2. Six forms of romanization are in use for Sanskrit. These are HK, IAST, Itrans, SLP1 and WX. Technology for easy conversion of Devanagari into any of these forms of Romanization is already available. It is therefore not necessary for the catalogue to store any metadata in Romanized form.

8. Subject classification for Sanskrit

- 8.1. Existing classification systems such as DDC, UDC and Library of Congress do not adequately classify the vast span of Sanskrit literature. Bajaj and Srinivas¹ have recently proposed a classification system for the traditional knowledge in India. The proposals embodied in this paper are indeed very suitable for the classification of Sanskrit literature but they need further practical validation, particularly with unpublished materials encountered in manuscripts.
- 8.2. It is imperative that a national standard is agreed and published on the classification of Indic literature quickly. The proposals of Bajaj and Srinivas need to be finalised as soon as possible. Once final they can be used for assigning subjects do the GBM manuscripts.

¹ Bajaj JK and Srinivas MD, Locating Indian knowledge in modern libraries: Incorporating the traditional classification of knowledge in India, Indian Journal of History of Science 2024 (59), 143-158

9. Cataloguing standards

- 9.1. The list of manuscript metadata required by Sanskrit scholars Is given at Annex-1.
- 9.2. A comparison of the metadata list given in Annex-1 against existing MARC21 and RDS standards is given at Annex-2. As can be seen from this comparison, there are number of metadata fields that are not covered by either of these standards.
- 9.3. It is therefore recommended that GBM adopt the list as per Annex-1 and use a text database such as the **open-source Mongo DB** to store the data. Search and delivery applications can be written for Mongo DB in all popular programming languages like C, C++, Rust, C#, Java, Node.js, Perl, PHP, Python, Ruby, Scala, Go and Erlang.
- 9.4. Where integration with other catalogues that follow MARC21 or other standards is required, data can easily be exported from the GBM database to the required format. Needless to say, the richness of the GBM database would not be available to such catalogues.

Annex-1

List of metadata² collected for each manuscript

1. Library details:
 - 1.1. *³Library / collection name
 - 1.2. *Library address / location
 - 1.3. *Library UID⁴
2. Manuscript identification
 - 2.1. Library manuscript ID⁵
 - 2.2. ID / locators in library catalogue (if any)
 - 2.3. Manuscript UID in GBM⁶
3. Provenance information
 - 3.1. Donor / source / provenance of manuscript
 - 3.2. Date manuscript was acquired by library
4. Manuscript Title information:
 - 4.1. Title found on manuscript
 - 4.2. Title given by the library in its catalogue

² The language and script in which each metadata field is to be recorded should be decided before any metadata is collected. An example could be that for Sanskrit language manuscripts in any script, the title field is to be entered in Devanagari script and Sanskrit language in prathamant

³ Metadata fields marked with an asterisk *, are to be taken from the list of authorities / lexica described later in this document.

⁴ Each library begiven auniquefouralphabet identifierin uppercaseRoman. The letters”O”and”I”should not be used asthey can confuse with the numbers 0 and 1. With 4 alphabets 3,31,776libraries / collections can be uniquely identified

⁵ Thisistheaccessionnumberassignedbythelibrarytoeachmanuscript.Ifacollectorhasnotassignedanaccessionnumber,theimagingteammustassignanaccessionnumbersequentiallybeginningwith1andrecord it on the datasheet placed with the manuscript.This datasheet should be the first image taken.

⁶ This is the UID accorded by GBM in its record. This should be unique and never repeated. A simple method that ensures uniqueness wouldbetousesthe libraryfouralphabetcode and the library’s accessionnumber as the GBMaccessionnumber.

- 4.3. *Title determined after reading the manuscript (including names of commentary / commentaries, etc.)
- 4.4. *Title UID(s)
- 4.5. For multi-text manuscripts⁷
 - 4.5.1. List of *titles of each text in manuscript (in the order of appearance in the manuscript)
 - 4.5.2. List of *title UIDs for each text
 - 4.5.3. Starting and end page number for each title
- 5. Subject information
 - 5.1. The *subject(s) applicable to the text in the manuscript
 - 5.2. *Subject UID
 - 5.3. List of *subjects for each text contained in a multi-text manuscript
- 6. Contents of the manuscript
- 7. Physical properties of the manuscript
 - 7.1. Substrate material(s) (e.g. paper, palm leaf etc.)
 - 7.2. Size length (to 0.5 cm precision)
 - 7.3. Size width (to 0.5 cm precision)
 - 7.4. Binding type
 - 7.5. Lines per page
 - 7.6. Letters per line
 - 7.7. Description of page layout
 - 7.8. Folio Information
 - 7.8.1. Folios numbered or not
 - 7.8.2. Number of folios
 - 7.8.3. List of missing pages
 - 7.8.4. Unrelated pages found
 - 7.8.5. Duplicated pages found

⁷A multi-text manuscript means a manuscript having more than one text is scribed in it (e.g. a collection of stotras; petank or sangraha).

- 7.8.6. Multiple pages of the same number. (due to a mistake in numbering)
- 7.8.7. Unreal missing pages due to a mistake in numbering, but the text is not missing
- 7.9. Number of images (pages)
- 7.10. Number of hands
- 7.11. Description of hands
- 7.12. Accented or not (Vedic manuscripts)
- 7.13. Overscribed or not
- 7.14. Illustrated or not
- 8. Condition of the manuscript
 - 8.1. Description of the condition
- 9. History of the manuscript
- 10. Identity and listing of important pages⁸
 - 10.1. Cover page
 - 10.2. First page of each text in the manuscript
 - 10.3. Last page of each text in the manuscript
 - 10.4. Colophons found
 - 10.5. Last page
 - 10.6. Flyleaf
 - 10.7. Illustrated pages
 - 10.8. Datasheet image
- 11. Important transcriptions from manuscript⁹
 - 11.1. Cover page
 - 11.2. Begin text / incipit (as written by scribe and in readable form; transcribed in the selected script, e.g. Devanagari for a Sanskrit

⁸ This requires each image file of each manuscript to have a unique file name. Unique page naming is essential and can easily be done by prefixing the library code and accession number to the page number of the manuscript.

⁹ All transcriptions must be exactly faithful to the manuscript. Scribal errors should not be corrected and unreadable text should be so identified.

- manuscript)
- 11.3. End text / explicit (as written by scribe and in readable form; transcribed in the selected script, e.g. Devanagari for a Sanskrit manuscript)
- 11.4. Flyleaf text
- 11.5. All colophons and rubrics
- 12. Date-related data (as obtained from manuscript)
 - 12.1. Year
 - 12.2. Year type (Shaka / Vikram, etc.)
 - 12.3. Samvatsar
 - 12.4. Season (ritu)
 - 12.5. Ayan
 - 12.6. Month
 - 12.7. Paksha
 - 12.8. Weekday (vaar)
 - 12.9. Date (tithi)
 - 12.10. Validity of the year
 - 12.11. Gregorian year (CE)
- 13. Person / place related information (as obtained from the manuscript)
 - 13.1. *Scribe(s)
 - 13.2. *Owner(s)
 - 13.3. *Accenter(s)
 - 13.4. *Editor(s)
 - 13.5. *Place(s) of writing
- 14. *Publication status on given date (e.g. published / probably unpublished)
 - 14.1. Publication UID
- 15. Description of the Digital images of the manuscript

- 15.1. File type
- 15.2. File size
- 15.3. Image resolution in ppi / dpi
- 15.4. Image cropped or not
- 15.5. Any text lost in cropping
- 15.6. Date of imaging
- 15.7. Imaging agency
- 15.8. Colour or Black & white images
- 16. Any additional point of significance about the manuscript
- 17. Persons and dates associated with the Manuscripts metadata entry
 - 17.1. Imaged by
 - 17.2. Date of imaging
 - 17.3. Image processed by
 - 17.4. Paginated by
 - 17.5. Pagination date
 - 17.6. Metadata recorded by
 - 17.7. Date of metadata entry
 - 17.8. Entry reviewed by
 - 17.9. Review date
 - 17.10. Uploaded to catalogue on

Comparison of metadata requirements from Annex-1 against existing standards

| S.No | Meta-data Element | Subfields | MARC21 | Dublin Core | BIB-FRAME |
|------|------------------------------|--|---|---|--|
| 1. | Library / Collection Details | Library/Collection Name; Library Address/Location; Library UID; Record No. | MARC21: 852 \$a (Location), 099 \$a (Call number), 035 \$a (System control no.) | DC: Coverage.Spatial; Identifier; Publisher | BIB-FRAME: bf:HeldBy; bf:Organization; bf:Local |
| 2. | Manuscript Title | Title (Main); Parallel Title; Uniform Title | MARC21: 245 \$a (Title), 246 (Variant title), 130 (Uniform title) | DC: Title | BIB-FRAME: bf:Title; bf:VariantTitle; bf:UniformTitle |
| 3. | Author / Creator | Personal Author; Corporate Author; Scribe; Attributed Author | MARC21: 100 \$a (Personal author), 110 (Corporate), 700 (Added entry) | DC: Creator | BIB-FRAME: bf:Person; bf:Organization; bf:Contribution |
| 4. | Physical Description | Extent (folios/pages); Dimensions; Material; Condition | MARC21: 300 \$a (Extent), 340 (Physical medium), 500 \$a (Condition note) | DC: Format; Description | BIB-FRAME: bf:Extent; bf:Material; bf:Physical-Condition |
| 5. | Language & Script | Language; Script; Secondary Language(s) | MARC21: 041 \$a (Language), 546 \$a (Language note) | DC: Language | BIB-FRAME: bf:Language; bf:Script |
| 6. | Date & Chronology | Date of Creation; Date Range; Calendar/Era | MARC21: 260 \$c (Publication date), 264 \$c (Production date), 518 (Date note) | DC: Date | BIB-FRAME: bf:Provision-Activity; bf:TemporalCoverage |

| | | | | | |
|-----|-------------------------------|---|---|-------------------------|--|
| 7. | Subject / Key-words | Topical Subjects; Names; Places; Events; Genres | MARC21: 650 \$a (Subject), 651 \$a (Geographic), 655 (Genre/form) | DC: Subject | BIB-FRAME: bf:Topic; bf:Place; bf:Genre-Form |
| 8. | Abstract / Summary | Summary; Scope; Content Note | MARC21: 520 \$a (Summary, abstract, annotation) | DC: Description | BIB-FRAME: bf:Summary |
| 9. | Provenance & Ownership | Provenance Note; Ownership Marks; Donor; Acquisition Details | MARC21: 561 \$a (Ownership/Provenance), 541 \$a (Immediate source of acquisition) | DC: Contributor; Source | BIB-FRAME: bf:Acquisition; bf:Ownership |
| 10. | Related Works / References | Related Titles; Citations; References | MARC21: 787 (Nonspecific relationship), 510 (Citation note) | DC: Relation; Source | BIB-FRAME: bf:Related-To; bf:BibliographicReference |
| 11. | Notes | General Notes; Custodial Notes; Binding Notes; Conservation Notes | MARC21: 500 (General note), 562 (Copy and version note) | DC: Description | BIB-FRAME: bf:Note |
| 12. | Identifiers | ISBN/ISSN; ISIL; Local Identifier; DOI/Handle | MARC21: 020 \$a (ISBN), 022 (ISSN), 035 (System control no.) | DC: Identifier | BIB-FRAME: bf:Identifier |
| 13. | Digital Access / Availability | Digital Object Identifier; URL; Access Conditions | MARC21: 856 \$u (URL), 506 (Restrictions on access) | DC: Identifier; Rights | BIB-FRAME: bf:ElectronicLocator; bf:Access-Policy |
| 14. | Rights & Permissions | Copyright Status; Usage Rights; Licensing | MARC21: 540 \$a (Terms governing use), 542 (Copyright status) | DC: Rights | BIB-FRAME: bf:UsageAndAccessPolicy; bf:CopyrightStatus |

| | | | | | |
|-----|------------------------------|---|--|-----------------------|---|
| 15. | Cataloging Details | Cataloger Name; Cataloging Date; Cataloging Rules | MARC21: 040 \$a (Cataloging source), 500 (Cataloging note) | DC: Contributor; Date | BIB-FRAME: bf:Admin-Metadata; bf:DescriptionConventions |
| 16. | Preservation / Conservation | Preservation Actions; Conservation Treatments; Storage Conditions | MARC21: 583 \$a (Action note), 852 \$c (Location details) | DC: Description | BIB-FRAME: bf:Preservation; bf:Condition |
| 17. | Relationships to Collections | Parent Collection; Series; Archival Fonds; Sub-collection | MARC21: 773 (Host item), 774 (Constituent unit), 775 (Other edition) | DC: Relation | BIB-FRAME: bf: Partof; bf: Collection |

Annex-3

List of Authorities / vocabulary lexica: Collections of unique information needed for various manuscripts

1. Information related to a Kruti / Rachana / Work
 - 1.1. Kruti UID
 - 1.2. Kruti Title, main and alternatives (e.g. Gitagovinda and Ashtapadi)
 - 1.3. Kruti UID
 - 1.4. Kruti Swaroop (Relationship of commentaries with main title. Mool, Tika, Bhashya, Translation, etc.) (can add ID)
 - 1.5. Parent of the Kruti UID (e.g. Mahabharata is the parent text for Bhagavat Gita) (Can have multiple Parents in upward multiple levels)
 - 1.6. Kruti Language(s) (can add ID)
 - 1.7. Kruti Chapter Info (Number and list of chapters)
 - 1.8. Kruti shloka count or equivalent
 - 1.9. Kruti form e.g. prose / poetry /both
 - 1.10. Transcription of Starting text of the Kruti
 - 1.11. Transcription of Ending text of the Kruti
 - 1.12. Transcription of Kruti colophon / rachnaprashasti
 - 1.13. *Subject of the kruti
 - 1.14. *Subject UID or List of subject UIDs
 - 1.15. *Keywords that can be applied to the contents of the Kruti
 - 1.16. Publication status (published / probably not published)
2. Information related to persons associated with a Kruti or a manuscript e.g. Author, commentator, Scribe, etc.
 - 2.1. Person UID
 - 2.2. Person name (without title of person like Acharya, Upadhyaya,

- etc.): main and aliases
- 2.3. Person titles e.g. Acharya, Upadhyaya, etc.
- 2.4. Persons' Gotra (can have unique ID for gotra)
- 2.5. Person's genealogy / vanshawali
- 2.6. Gender
- 2.7. Period when person lived
- 2.8. Places associated with person
- 2.9. Patrons / patronage
- 2.10. Preceptor(s) / guru(s) including UIDs
- 3. Gotra Information
 - 3.1. Gotra UID
 - 3.2. Gotra name
 - 3.3. Alternative names for the Gotra
 - 3.4. Gotra genealogy / vanshawali
- 4. Place Information (where the manuscript was scribed or the Kruti was created)
 - 4.1. Place UID
 - 4.2. Name of Place including alternative names (e.g. Kashi, Varanasi etc.)
 - 4.3. Name of Rajya
 - 4.4. Description of the place where available
 - 4.5. Brief history of the place
- 5. Subject information
 - 5.1. Subject UID
 - 5.2. Name of the subject

- 5.3. Alternative subject names for each subject if applicable
- 5.4. Parent of the subject
- 5.5. UID of parent

- 6. Library / collection information
 - 6.1. Library UID
 - 6.2. Full name of library
 - 6.3. Full postal address of the library
 - 6.4. GPS co-ordinates of the library
 - 6.5. State
 - 6.6. Details of the library head
 - 6.6.1. Name
 - 6.6.2. Job title
 - 6.6.3. Email ID
 - 6.6.4. Contact phone number
 - 6.7. Parent organization if applicable
 - 6.8. Details of the head of the parent organisation
 - 6.8.1. Name
 - 6.8.2. Job title
 - 6.8.3. Email ID
 - 6.8.4. Contact phone number
 - 6.9. Brief description of the work of the institution
 - 6.10. Digital copy of the MoU with the library
 - 6.11. Whether library permission needed before download of manuscripts
 - 6.12. Pricing regime
 - 6.13. Payment bank details
 - 6.14. Authoritative list of sources / manuscript provenance



Working Group III

Theme: Manuscriptology and Palaeography

Coordinator

Prof. Sudhir Kumar Lall

HoD (Kalakosa Division), IGNCA

Co-Coordinator

Ms. Pooja Hali

Director, Ministry of Culture, Govt. of India

Experts

Prof. Vasantkumar M. Bhatt

Formerly Director of the School of Languages, Gujarat University, Ahmedabad

Dr. Sweta Prajapati

I/c Director, Oriental Institute, Maharaja Sayajirao University, Baroda

Shri P. L. Shaji

Senior Manuscript Conservator, ORIML, University of Kerala

Dr. T. S. Ravishankar

Former Director (Epigraphy), ASI, Mysore

Prof. Akhlaque Ahmad

Professor, Centre of Persian and Central Asian Studies, JNU

Dr. Kamal Kishor Mishra

HoD & Associate Professor, Dept. of Sanskrit, University of Kolkata

Dr. Abhijit Dixit

Regional Director, R C Varanasi, IGNCA

Revitalising India's Manuscript Heritage

A Report ¹

The Legacy of India's Manuscript Heritage

India's manuscript heritage represents one of the most extensive and diverse textual traditions in the world. With over 5.2 million manuscripts documented by the National Mission for Manuscripts (NMM)—and many more believed to reside in private collections, temple libraries, and regional archives—the scale of this legacy is extraordinary. These manuscripts span more than 80 languages and are written in dozens of scripts, including Brahmi, Sharada, Grantha, Nandinagari, Kaithi, Modi, and Persian forms such as Naskh and Nastaliq.

This corpus is distinguished not only by its volume but by its civilizational uniqueness. It encompasses a vast range of disciplines—metaphysics, astronomy, medicine, jurisprudence, poetics, ritual theory, and regional knowledge systems—reflecting centuries of intellectual inquiry and cultural transmission, which have been a hallmark of Indian traditions. India's manuscript tradition is distinguished not only by its intellectual breadth but also by the remarkable diversity of its writing surfaces—each rooted in regional ecology, artisanal ingenuity, and cultural aesthetics. India's manuscript tradition is distinguished not only by its intellectual breadth but by the remarkable diversity of its writing surfaces—each rooted in regional ecology, artisanal ingenuity, and cultural aesthetics. In the humid tropics of South India, talipat palm leaves (*Corypha umbraculifera*) were abundant and ideal for stylus-based inscription. In contrast, in the temperate zones of Kashmir, such palms were largely unavailable, prompting the use of *bhurja patra* (*Betula utilis*), a resilient birch bark suited to the region's climate. In Assam, *sanchipat*—prepared from the bark of *Aquillaria agallocha*—offered a durable and visually rich medium, especially within Vaishnavite monastic

1. Coordinator- Prof. Sudhir Lall, Co-Coordinator-Ms. Pooja Hali, Members -Prof. Vasantkumar M. Bhatt, Dr. Sweta Prajapati, Shri P. L. Shaji, Dr. TS Ravishankar, Dr. Kamal Kishore Mishra, Dr. Abhijit Dixit, Dr. Sarwarul Haq.

traditions. Elsewhere, materials like handmade paper in Rajasthan, cloth scrolls in Gujarat, and copper plates in Tamil Nadu, Maharashtra, and Gujarat reflect localized innovations in textual preservation. Odisha, though more commonly associated with stone inscriptions and temple architecture, also holds a lesser-known corpus of copper plate grants issued by dynasties such as the Bhauma-Kara and Somavamshi. Yet, as ecological pressures mount and traditional harvesting practices decline, some of these materials have become unsuitable for long-term conservation or unsustainable for continued use—necessitating thoughtful stewardship and adaptive preservation strategies. These writing surfaces are not merely functional—they are cultural artifacts that encode the aesthetic, ritual, and philosophical sensibilities of their regions, making each manuscript a unique embodiment of civilizational memory. Unlike centralized manuscript traditions elsewhere, India's textual culture is pluralistic and decentralized, shaped by contributions from monastic institutions, royal courts, local scholarship, merchant guilds, and community custodians. This pluralistic culture of textual stewardship transforms the Indian manuscript archive into a dynamic repository of inquiry and a living expression of knowledge as a civilizational ethic.

Despite its richness, manuscript-related activities—such as cataloguing, documentation, digitization, conservation (both preventive and curative), and dissemination—require a holistic framework. A coordinated national effort is essential to safeguard this unique heritage meaningfully, mitigate its vulnerability to environmental and institutional neglect, and revitalize its enduring legacy. For this purpose, a multi-pronged and deeply coordinated approach is essential. This begins with the systematic documentation and cataloguing of manuscripts using harmonized metadata standards that ensure consistency, discoverability, and long-term interoperability. Fragile materials must be digitized and preserved with sensitivity to their physical condition and script-specific complexities, including the development of custom OCR (Optical Character Recognition) and HTR (Handwritten Text Recognition) models that honour the nuances of regional and classical scripts. Equally vital is the cultivation of human expertise; scholars and technicians must be rigorously trained in palaeography, conservation science, and editorial protocols that strike a balance between precision and cultural sensitivity. Manuscript studies should be integrated into higher education curricula and made accessible through public knowledge platforms, bridging academic inquiry with civic engagement. Ultimately, the effort must be grounded in collaborative frameworks that span institutions, languages, and disciplines—

creating a dynamic ecosystem of stewardship, scholarship, and shared responsibility.

II A Brief Survey of Existing Manuscript-Related National Initiatives in India

Preserving India's vast and unparalleled manuscript heritage as her civilizational archive is an uphill task, and several institutions of national importance are relentlessly working in this direction. A brief overview of the Indira Gandhi National Centre for the Arts (IGNCA), the National Mission for Manuscripts, select Sanskrit universities and Oriental Research Institutes (ORI), and the French Institute of Pondicherry is given below --

1. Indira Gandhi National Centre for the Arts (IGNCA)

Established in 1987 under the Ministry of Culture, IGNCA is a premier interdisciplinary institution dedicated to the study and preservation of Indian arts and knowledge systems. It has served as the original nodal agency for the National Mission for Manuscripts (NMM) and continues to play a vital role in the digitization of manuscripts and cultural archiving.

Key Contributions:

- **Manuscript Digitization and Metadata:** IGNCA has developed standards for cataloguing and metadata encoding, using Unicode and multilingual search capabilities. Its digitization efforts prioritize fragile and rare manuscripts, often sourced from private collections and regional repositories.
- **Kalasampada Digital Repository:** IGNCA's Kalasampada platform is a multimedia digital library that integrates manuscripts, rare books, photographs, audio-visual materials, and scholarly publications. It has digitized over 12,000 microfilm rolls and 100,000 slides, including materials from institutions such as the British Library and the Victoria & Albert Museum. The facility is currently available only on the intranet, due to issues related to pricing, possession, Intellectual Property Rights, and copyright, among others. Although partial information can be accessed from the IGNCA's website at ignca.gov.in.
- **Training and Outreach:** IGNCA conducts workshops in manuscriptology, conservation, and digital archiving, often in collaboration with universities and cultural bodies.
- **Philosophical Integration:** True to its founding vision, IGNCA emphasizes

the aesthetic, metaphysical, and performative dimensions of India's textual traditions, treating manuscripts not merely as data but as living cultural expressions.

IGNCA's approach is holistic, bridging textual, visual, and oral traditions, and it remains central to India's cultural preservation infrastructure.

2. National Mission for Manuscripts

The National Mission for Manuscripts (NMM), launched in 2003 by the Government of India, was tasked with identifying, documenting, conserving, and making accessible India's vast manuscript heritage. Conceived as a national initiative, it served as a timely, transformative, and necessary response to the urgent challenge of reclaiming a civilizational inheritance—much of which remained in fragile and deteriorating condition. As a comprehensive, pan-Indian effort, the Mission addressed both the conservation of manuscripts and the dissemination of the knowledge they embodied. It stood not only as a repository of textual wisdom but as a catalyst for cultural renewal and scholarly engagement.

Core Objectives:

- **Survey and Documentation:** Over 5.2 million manuscripts have been documented across India. The Mission maintained a national database using Dublin Core metadata standards.
- **Digitization and Access:** More than 3.5 lakh manuscripts (covering 3.5 crore folios) were digitized, with 76,000 manuscripts available for public access via its website (namami.gov.in).
- **Conservation and Restoration:** Through Manuscript Conservation Centres (MCCs), the Mission supported preventive and curative conservation using scientific methods.
- **Research and Publication:** It brought out critical editions, translations, and facsimile publications of rare texts, including Sanskrit, Persian, and Arabic manuscripts.
- **Capacity Building:** Training programs in manuscriptology, paleography, and conservation were conducted nationwide, often in partnership with Sanskrit universities and Oriental Institutes.

3. Sanskrit Universities

India hosts over 18 dedicated Sanskrit universities, each committed to the

revival, teaching, and research of Sanskrit and allied disciplines. These institutions serve as intellectual anchors for manuscript studies and traditional knowledge systems. Some notable universities are enumerated below, which have significant and active connections with the theme under review:

University Location Highlights

Sampurnanand Sanskrit Vishvavidyalaya Varanasi, UP Oldest Sanskrit university; rich manuscript collection and editorial tradition.

- Central Sanskrit University, New Delhi Offers distance education, research fellowships, and digitization projects.
- Shri Lal Bahadur Shastri National Sanskrit University, New Delhi Focuses on teacher training and Shastric research.
- Kameshwar Singh Darbhanga Sanskrit University Bihar Known for its manuscript repository and regional outreach.
- Sree Sankaracharya University of Sanskrit, Kalady Kerala Integrates Sanskrit with modern disciplines and cultural studies.

Table:01

Contributions:

- Curriculum Integration: Courses in manuscriptology, palaeography, and textual criticism are increasingly part of postgraduate programs.
- Research and Publication: Universities publish critical editions, commentaries, and translations of rare manuscripts.
- Collaborations with NMM: Many serve as MRCs or MCCs under the national mission, contributing to survey and conservation efforts.
- Digital Initiatives: Institutions like Madras Sanskrit College have launched virtual campuses and online Sanskrit learning platforms.

These universities are not merely academic institutions—they are custodians of India's intellectual traditions, nurturing a new generation of scholars and manuscriptologists.

4. Oriental Research and other Institutes

Oriental Institutes represent a legacy of scholarship of Indic Studies dating back to the colonial and early post-colonial periods. Their focus on editing, publishing, and preserving Sanskrit and Prakrit manuscripts has been

foundational to modern manuscriptology. Some prominent institutes under this umbrella are as follows:

- Bhandarkar Oriental Research Institute (BORI), Pune: Home to the Rigveda, Natyashastra, and Srimad Bhagavat Gita manuscripts listed in UNESCO's Memory of the World Register.
- Sarasvati Mahal Library, Thanjavur: One of the oldest libraries; rich in palm-leaf manuscripts and royal archives.
- Adyar Library and Research Centre, Chennai: Houses over 18,000 manuscripts and rare Indological texts.
- Kuppuswami Sastri Research Institute, Chennai: Known for critical editions and shastraic research.
- Asiatic Society, Kolkata: Founded in 1784, the Society's manuscript collection is vast and multilingual, spanning Indian, Asian, and Western languages. Organized into Sanskrit, Islamic, Sino-Tibetan, and English sections, it features palm-leaf and paper manuscripts, rare illustrated texts, Buddhist xylographs, imperial holdings, and early European works—reflecting a rich tapestry of global intellectual and cultural heritage. It prides itself on manuscripts of the Padapatha of the Rigveda, and the Shahnama, etc.
- Mythic Society, Bengaluru: The repository has over 12,000 manuscripts, with over 1.2 million folios about the Indian Knowledge System. Some very important Indian manuscripts, such as the Arthashastra, Lilavati, and Bijaganita, figure in its collection.
- Rajasthan Oriental Research Institute, Jodhpur: The Institute's total manuscript collection, comprising approximately 1.24 lakh manuscripts, is housed across its headquarters and six branch repositories in Alwar, Bikaner, Jaipur, Bharatpur, Chittorgarh, Udaipur, and Kota.
- French Institute, Pondicherry: The Department of Indology at this institute advances the study of Indian heritage, with emphasis on South Indian traditions. Through manuscript stewardship, linguistic research, and interdisciplinary collaboration, it fosters global scholarship. Key initiatives include the Shaiva siddhanta manuscript collections, which have been recognised by UNESCO under the Memory of the World Program. It is also working on and pioneering material profiling of palm-leaf texts from Tamil Nadu and beyond. It also has an active major research

project - the ‘Palm-Leaf Manuscript Profiling Initiative’ (PLMPI), which intends to carry out innovative and comprehensive material studies of approximately 300 palm-leaf manuscripts from the private collections of the French Institute of Pondicherry (IFP) and the Pondicherry Centre of the Ecole Française d’Extrême-Orient (EFEO).

- **Oriental Research Institute and Manuscripts Library, Thiruvananthapuram:** The Oriental Research Institute and Manuscripts Library of the University of Kerala shelter a vast and historically rich collection: approximately 30,000 codices or bundles containing around 60,000 individual works, inscribed on palm leaf, paper, copper plates, and bark, and spanning a wide array of scripts and languages. ORI Trivandrum is renowned for its landmark publication series—the Trivandrum Sanskrit Series—which unearthed and critically edited rare and previously unpublished Sanskrit texts. Under the editorship of Mahamahopadhyaya T. Ganapati Sastri, the institute brought to light the lost plays of Bhasa, including *Svapnavasavadattam*, *Pratijnayaugandharayanam*, and *Avimraka*, marking a watershed moment in modern Indic studies.

Key Activities:

- **Editing and Publication:** Institutes publish critical editions with commentaries, often in bilingual formats.
- **Preservation and Conservation:** Many have adopted modern conservation techniques and collaborate with NMM for digitization.
- **Scholarly Networks:** These institutes host seminars, fellowships, and collaborative research projects, often engaging with international scholars.
- **Philosophical Depth:** Their work reflects a commitment to the metaphysical and ethical dimensions of Indian knowledge systems.

Toward Integration and Future Directions:

While each initiative operates with distinct mandates, there is growing recognition of the need for integration—technically, institutionally, and philosophically. Efforts like metadata harmonization can be beneficial, as India moves toward a more integrated and digitally empowered future, where these institutions will be well-positioned to play a crucial role in shaping the contours of manuscriptology, cultural memory, and civilizational renewal.

III. Scripts of India: Diversity and Historical Trajectories

The Indian subcontinent is home to one of the most diverse and historically layered script traditions in the world. From the earliest inscriptions in Brahmi to the flourishing of regional scripts and the later introduction of Persian-Arabic calligraphic styles, the evolution of writing systems in India reflects a complex interplay of linguistic, religious, political, and artistic forces. This overview outlines the major trajectories of script development, with a focus on Brahmi-based scripts, regional variants, and the integration of Perso-Arabic scripts during the medieval period.

I. Brahmi: The Foundational Script

The Brahmi script, attested from the 3rd century BCE in Ashokan inscriptions, is widely regarded as the progenitor of most Indic scripts. Its origins remain debated—some scholars propose a derivation from Semitic scripts via Aramaic, while others argue for indigenous development. Regardless of its genesis, Brahmi's structural innovation lay in its syllabic nature, representing consonant-vowel units (akṣaras), and its capacity to encode the phonetic richness of Indo-Aryan languages.

Key Features:

- Written left to right
- Consonant characters with inherent vowel /a/
- Diacritics to modify vowels and consonants
- Use in Prakrit inscriptions and Buddhist texts

Over time, Brahmi diversified into regional styles, influenced by local calligraphic preferences and linguistic needs.

II. Regional Evolution of Brahmi-Derived Scripts

1. Northern Zone: Gupta, Sharada, and Nagari

- Gupta Script (4th–6th century CE): A transitional form between Brahmi and later northern scripts, used in Sanskrit inscriptions and manuscripts. It introduced more cursive and rounded forms.
- Sharada Script (from 3rd century CE onwards): Developed in Kashmir and used for Sanskrit and Kashmiri texts. It is distinguished by its angular elegance and structural clarity, tailored for the precise rendering of Sanskrit in Kashmir and the northwestern regions. Its lineage reflects a

unique transitional phase between Gupta Brahmi and later northern scripts, preserving archaic features while enabling scholastic transmission.

- Nagari and Devanagari: By the 10th century, Nagari emerged as a dominant script in northern India. Devanagari, its standardized form, became the principal script for Sanskrit, Hindi, Marathi, and Nepali. It features a horizontal head-stroke (shirorekha) and systematic vowel notation.

2. Eastern Zone: Siddham, Gaudi, Bengali-Assamese

- Siddham: Used for Buddhist texts, especially in East Asia. It evolved from the Gupta and had an influence on Japanese esoteric Buddhism.
- Gaudi: A transitional script in Bengal and Odisha, giving rise to:
- Bengali-Assamese Script: Characterized by its looped forms and absence of the head-stroke. Used for Bengali, Assamese, and Meitei.
- Odia Script: Developed rounded forms suited to palm-leaf manuscripts, with distinct vowel markers.

3. Southern Zone: Tamil, Grantha, Telugu-Kannada, Malayalam

- Tamil Script: Derived from southern Brahmi, it underwent simplification and standardization. It avoids complex consonant clusters, reflecting the phonology of Tamil.
- Grantha Script: Used in Tamil Nadu for writing Sanskrit. It preserves complex conjuncts and is still used in temple manuscripts.
- Telugu-Kannada Scripts: These sister scripts evolved from the Kadamba script. The Telugu script is more rounded, while the Kannada script retains angular features.
- Malayalam Script: Derived from Grantha and Vatteluttu, it was adapted to the phonetics of Malayalam and palm-leaf writing.

4. Western Zone: Modi and Mahajani

- Modi Script: Used in Maharashtra for administrative purposes. It is cursive and suited for rapid writing.
- Mahajani and Kaithi: Commercial scripts used in Rajasthan, Bihar, and Uttar Pradesh. They were employed for account-keeping and informal correspondence.

III. Non-Brahmi Scripts and their localised Variants

- While Brahmi-derived scripts dominate the subcontinent, several non-Brahmi scripts have flourished:
- Kharosthi: Used in Gandhara (northwest India) from the 3rd century BCE to the 3rd century CE. Derived from Aramaic and written right to left. Used for Prakrit and Buddhist texts.
- Burmese, Thai, and Khmer Scripts: Though outside India, these Southeast Asian scripts evolved from Pallava and Grantha models, showing India's transregional influence.
- Tribal Scripts: In modern times, indigenous communities have developed scripts such as:
 - Ol Chiki (Santali)
 - Warang Citi (Ho)
 - Sorang Sompeng (Sora). These scripts reflect cultural assertion and linguistic identity.

IV. Persian-Arabic Script Traditions in India

With the advent of Islamic rule in the 12th century, Persian and Arabic scripts were introduced to the Indian subcontinent, bringing a new administrative and aesthetic dimension. The Naskh and Nastaliq scripts played a pivotal role in shaping the visual culture of Arabic, Persian, and Urdu texts. Naskh, with its clear and rounded contours, was suited to Quranic manuscripts and administrative records, while Nastaliq—renowned for its elegant, cascading form—became the preferred style for Persian poetry and Indo-Persian literary expression, especially under Mughal patronage. Urdu adopted the Nastaliq script, modifying it to accommodate the phonetic nuances of the Indic language. Written right to left, it features intricate ligatures and context-sensitive letter forms, and emerged as a vibrant medium for poetry, prose, and journalism across northern India. During this era, multilingual manuscripts were also en vogue. Many manuscripts from the medieval period feature bilingual or trilingual scripts—e.g., Persian in Nastaliq alongside Sanskrit in Devanagari or Grantha. This reflects India's layered linguistic and cultural milieu.

V. Colonial and Postcolonial Developments

1. Standardization and Print

- The colonial period saw the standardization of scripts for printing—especially Devanagari, Bengali, and Tamil.
- Movable type and lithography influenced script forms and orthographic conventions.

2. Scripts as Identity

- Discussions around script preferences—such as Hindi in Devanagari and Urdu in Nastaliq—gained cultural and political significance during the nationalist period, reflecting broader questions of identity, representation, and linguistic heritage.
- Post-independence, regional scripts were promoted through education and media, though some commercial scripts like Modi declined.

3. Digital Encoding and Unicode

- The advent of Unicode has enabled the digital representation of most Indian scripts.
- Projects like the TDIL (Technology Development for Indian Languages) and Google’s Noto fonts have expanded script accessibility. The latter offers a unified, open-source typeface family that supports over 1,000 languages and 150 writing systems. Designed to eliminate “tofu” (blank boxes for unsupported characters), Noto ensures consistent, legible typography across global scripts, making it indispensable for multilingual publishing, digital preservation, and cross-cultural communication.
- However, challenges remain in encoding complex conjuncts and rare manuscript variants.

The evolution of scripts in India is not merely a technical history—it is a story of cultural transmission, aesthetic innovation, and linguistic plurality. From the geometric clarity of Brahmi to the ornate elegance of Nastaliq, and from palm-leaf manuscripts to digital archives, Indian scripts embody the civilizational ethos of continuity and adaptation. Institutionalising script studies—through palaeography labs, comparative atlases, and digital tools—can ensure that this legacy remains vibrant and accessible for future generations.

IV Deciphering Indian Scripts: Palaeographic Tools and Techniques

Indian palaeography, the study of ancient scripts and writing systems, is a discipline that bridges the technical precision of decipherment with the

philosophical depth of textual transmission. It is much more than merely a forensic exercise in reading old letters—it is a civilizational dialogue with the past, requiring a confluence of linguistic insight, historical context, and material awareness. The decipherment of Indian scripts, ranging from Brahmi and Kharosthi to regional variants such as Grantha, Modi, and Sharada, requires a nuanced toolkit that integrates traditional scholarship with evolving digital methodologies.

Historical Foundations and Script Evolution

At the heart of Indian palaeography lies the recognition of script as a dynamic cultural artifact. The evolution from Brahmi (3rd century BCE) to its regional offshoots reflects not only phonetic adaptations but also aesthetic and political shifts. Palaeographers must first situate a manuscript within its historical and geographical milieu—identifying the script family, chronological strata, and linguistic register. This foundational step is crucial for distinguishing between orthographic conventions and scribal idiosyncrasies.

Core Tools for Decipherment

Decipherment begins with the careful observation of letterforms, ligatures, and diacritical marks. Traditional tools include:

- **Script Charts and Comparative Tables:** These provide visual mappings of character evolution across time and regions. For example, tracing the transformation of the Brahmi “ka” into its Nagari, Bengali, or Tamil variants.
- **Ink and Writing Material Analysis:** Understanding the medium—whether palm leaf, birch bark, or paper—helps contextualize the script’s morphology. Stylus-based scripts (e.g., Grantha on palm leaf) often exhibit angular forms, while pen-based scripts (e.g., Devanagari on paper) allow for more fluid strokes.
- **Orthographic Conventions:** Familiarity with scribal abbreviations, punctuation, and numeral systems is essential. For instance, the use of “sandhi” ligatures or “virama” marks in Indic scripts can dramatically alter phonetic interpretation.

Methodologies of Decipherment

Indian palaeography employs a layered methodology that combines visual scrutiny with linguistic reconstruction:

1. Letterform Analysis and Pattern Recognition

This involves identifying recurring shapes and their contextual variants. Palaeographers often begin with high-frequency characters (e.g., vowels or common consonants) and build a reference set. Pattern recognition is especially vital in cursive or stylized scripts, such as Modi or Nandi Nagari, where characters may be compressed or conjoined.

2. Phonetic Reconstruction

Once basic graphemes are identified, scholars reconstruct phonetic values using known linguistic rules. This is particularly challenging in transitional scripts or bilingual manuscripts, where phonetic drift may occur. Comparative linguistics—drawing parallels with languages such as Sanskrit, Prakrit, Tamil, or Persian—often aids this process.

3. Contextual and Semantic Validation

Decipherment is not complete until the reconstructed text makes semantic sense. This requires familiarity with genre conventions (e.g., colophons, invocatory verses, calendrical formulas) and cultural idioms. For example, identifying a ‘shloka’ meter or a ‘tithi’ reference can validate the accuracy of transcription.

4. Cross-Script Comparison

Many Indian manuscripts exhibit multilingual or multiscript features—such as Sanskrit in Grantha alongside Tamil in Vatteluttu. Palaeographers must be adept at switching registers and recognizing inter-script borrowings. This comparative agility is especially crucial in Indo-Persian manuscripts or Buddhist texts with Tibetan glosses.

Digital and Computational Enhancements

Recent advances have introduced digital tools that augment traditional palaeographic methods:

- **High-Resolution Imaging and Multispectral Analysis:** These techniques reveal faded or overwritten text, especially on fragile palm leaves or water-damaged folios.
- **HTR (Handwritten Text Recognition) Models:** Custom-trained models for Indic scripts—though still nascent—are being developed to automate

transcription. Projects like Transkribus and eScriptorium are exploring Indic integrations.

- **Metadata Encoding and TEI Standards:** Encoding palaeographic features using XML-based standards allows for interoperable digital editions and scholarly collaboration.

Challenges and Philosophical Reflections

The interpretive gaps in decipherment serve as productive sites for inquiry, enabling the refinement of readings and the advancement of understanding. Scribal errors, regional variants, and damaged folios often leave gaps that require interpretive judgment. In such contexts, the palaeographer assumes the role of a discerning steward—advancing decipherment through a calibrated balance of fidelity and informed interpretation. Moreover, Indian palaeography invites a philosophical reflection on the nature of transmission. Each script and glyph constitutes a material trace of the intellectual tradition, marking the intersection of thought, practice, and cultural transmission. The act of decipherment is thus an act of revival, of reanimating voices long silent.

Toward a Living Palaeography

To revitalize Indian palaeography, there is a need for:

- **Interdisciplinary Training:** Combining philology, material science, and digital humanities.
- **Institutional Collaboration:** Linking repositories, scholars, and technologists across India and globally.
- **Public Engagement:** Making deciphered texts accessible through translations, exhibitions, and educational platforms.

V. Textual Criticism: Reinstating Linguistic and Metrical Foundations in Indian Manuscriptology

Textual criticism is the scholarly pursuit of reconstructing a lost original—termed the archetype—through the analysis of manuscript variants and transmission histories. It is considered a branch of linguistics and reference bibliography, relying on philological insight and human judgment rather than a fixed methodology. This is where the disciplines of manuscriptology and palaeography converge to re-create a ‘text’ in its cultural and civilizational context. Textual criticism as an academic discipline in India demands renewed attention as both a scholarly and institutional imperative. Traditionally

viewed as a technical exercise, it holds deeper epistemological significance, especially within Sanskrit studies. While Western frameworks offer pluralistic methodologies, Indic traditions have long practiced rigorous textual transmission and emendation, often guided by commentarial traditions and internal consistency rather than formalized critical apparatus. The absence of structured training in textual criticism across Indian academic programs has led to a reliance on received texts, undermining scholarly rigor. Bridging these traditions requires curricular reform and public engagement. Introducing dedicated modules in postgraduate programs and collaborating with global institutions can foster methodological awareness. Furthermore, digital tools tailored for Indian scripts present new opportunities for textual analysis. A policy shift is essential to reposition textual criticism as a foundational discipline, integrating it into mainstream pedagogy and research. Synthesizing Western and Indic approaches can revitalize Sanskrit scholarship and reaffirm its relevance within contemporary academic discourse.

Simultaneously, three core competencies are indispensable for any critical text editor working within the domain of Indian manuscriptology:

- a. Linguistic mastery—proficiency in Sanskrit, Pali, Prakrit, Apabhramsha, and the regional languages of the medieval period;
- b. Subject-matter depth—a foundational and nuanced understanding of the intellectual and cultural content embedded in the manuscripts;
- c. Metrical literacy—in cases where the manuscript is composed in verse, a working command of Sanskrit and Prakrit prosody is essential.

These prerequisites form the bedrock of any serious engagement with manuscript traditions. This legacy is exemplified by centres such as the Indira Gandhi National Centre for the Arts, Bhandarkar Oriental Research Institute, Maharaja Sayajirao Gaekwad Oriental Research Institute, and the French Institute (Pondicherry), whose seminal contributions to text-critical editions affirm the vitality of Indian engagement with textual criticism. Strengthening academic programs and policy frameworks must now align with and amplify such efforts.

VI. Manuscript Pedagogy in India: Structures and Shifts

The study of manuscripts in India underscores a civilizational responsibility—one that extends beyond technical engagement to preserve the continuity of cultural memory and intellectual heritage. As the country undertakes large-scale digitization and conservation of its textual heritage, the need for trained

scholars in manuscriptology and palaeography has become increasingly urgent. Over the past two decades, a growing number of academic institutions have responded by offering structured programs that blend traditional philology with modern archival science. These programs vary in depth, scope, and orientation, but together they form a nascent ecosystem of manuscript education in India.

At the heart of this movement is perhaps a perception that manuscripts are not static remnants but dynamic carriers of knowledge. Their decipherment, preservation, and interpretation require a multidisciplinary approach—one that draws from linguistics, history, manuscriptology, conservation science, and digital humanities. Accordingly, the syllabi of manuscriptology programs across India reflect a layered pedagogy: from script training and textual criticism to metadata encoding and conservation techniques.

Institutional Anchors and Programmatic Diversity

- The most comprehensive and nationally visible program is offered by the Indira Gandhi National Centre for the Arts (IGNCA), New Delhi. Its Postgraduate Diploma in Manuscriptology and Palaeography is structured around eight modules, covering everything from the genealogy of Indian scripts to hands-on training in Brahmi, Sharada, Grantha, Sharada, Modi, Takari, and Nastaliq. The program emphasizes both theoretical grounding and practical engagement, with students expected to edit actual manuscripts and present their findings through a viva voce presentation. IGNCA's curriculum is notable for its integration of conservation science and digital tools, including imaging protocols and metadata standards.
- In southern India, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya (SCSVMV) in Kanchipuram offers a full-fledged academic department dedicated to manuscriptology. Its programs span undergraduate, postgraduate, and doctoral levels, with specialized training in regional scripts such as Vattezhuthu and Nandinagari. The university's collaboration with MeitY's Saswathaiswaryam project has led to the development of metadata standards for palm-leaf manuscripts, positioning it as a leader in script-specific archival research. SCSVMV also integrates museology and epigraphy into its curriculum, reflecting a broader orientation in heritage studies.
- Kerala's ORIMSS (Oriental Research Institute and Manuscript Studies)

offers a Master's program tailored to the state's rich manuscript traditions. Students are trained in deciphering Brahmi, Grantha, and Vattezhuthu, and are introduced to both traditional conservation methods and modern digitization techniques. The program includes fieldwork components, allowing students to engage directly with manuscript repositories and local custodians.

- Beyond these flagship institutions, a wide array of universities offer certificate, diploma, and postgraduate programs in manuscriptology. These include Jawaharlal Nehru University (New Delhi), the University of Pune, Tamil University (Thanjavur), Osmania University (Hyderabad), the University of Mysore, the University of Madras, and Rashtriya Sanskrit Vidyapeeth (Tirupati), among others. While the depth and specialization of these programs vary, most include core modules on palaeography, codicology, cataloguing, and textual editing.

Curricular Themes and Pedagogical Structures

Despite institutional diversity, certain curricular themes recur across programs. Introductory modules typically cover the definition and scope of manuscriptology, types of manuscripts, and their historical development. Palaeography modules focus on the evolution of Indian scripts, with emphasis on Brahmi, Kharosthi, Gupta, Nagari, and regional variants. Students learn to identify letterforms, ligatures, and scribal conventions, often through comparative script charts and facsimile analysis.

Codicology—the study of manuscript materials and formats—is another core component. Students examine the physical characteristics of manuscripts, including writing surfaces (such as palm leaf, birch bark, and paper), ink composition, stylus techniques, and foliation systems. This material awareness is crucial for conservation and digitization, as it informs imaging protocols and restoration strategies.

Cataloguing and metadata encoding are increasingly emphasized, especially in programs aligned with national digitization efforts. Students are introduced to classification systems, Dublin Core standards, and XML-based encoding frameworks such as TEI. This technical training enables graduates to contribute to digital repositories and interoperable archival platforms, thereby enhancing their ability to support digital preservation.

Script training remains a cornerstone of manuscript education. Institutions

offer instruction in a wide range of scripts, including Grantha, Sharada, Modi, Nandi Nagari, Takari, Newari, and Nastaliq, as mentioned above. Some programs also include Tibetan, Persian, and Southeast Asian scripts, reflecting India's multilingual manuscript heritage. Training typically involves both reading and writing exercises, supported by facsimile reproductions and transcription assignments.

Textual criticism and editing form the final layer of most syllabi. Students learn to identify textual variants, reconstruct damaged passages, and prepare critical editions. This involves familiarity with prosody, grammar, and genre conventions, as well as philosophical sensitivity to the text's intended meaning. Editing is often paired with translation and commentary, encouraging students to engage with manuscripts as living texts rather than inert objects.

Emerging Trends and Challenges

One of the most promising trends is the integration of digital humanities into manuscript education. Institutions are beginning to incorporate tools such as IIIF (International Image Interoperability Framework), OCR (Optical Character Recognition), and HTR (Handwritten Text Recognition) for Indic scripts. While these technologies are still in development, their inclusion in syllabi signals a forward-looking approach to manuscript studies.

Another trend is the expansion of public engagement. Certificate courses, workshops, and exhibitions are being used to democratize manuscript knowledge and attract non-specialist audiences. This outreach is particularly important for building awareness and support for manuscript preservation.

However, challenges remain. There is no standardized national curriculum for manuscriptology, leading to uneven training and limited scholarly mobility. Script training is often constrained by faculty availability and resource limitations. Digital integration, although growing, remains patchy and reliant on institutional partnerships. Moreover, many programs lack sustained funding and infrastructural support, limiting their long-term impact.

As India invests in digitization and manuscript preservation, it must also invest in decoding capacity—through training, infrastructure, and interdisciplinary research. Only then can the silent voices of its manuscripts be heard again, not as remnants of the past, but as living texts.

VII. Global Best Practices in Manuscript Digitisation and Palaeography

This section presents a comparative analysis of manuscript digitization and palaeographic initiatives across four regions—the European Union (EU), the United Kingdom (UK), the United States (US), and Southeast Asia (SEA). Drawing on verified institutional practices and scholarly frameworks, it examines technical standards, metadata protocols, access models, and cultural approaches. The study highlights how each region balances conservation, accessibility, and scholarly engagement, offering insights for future manuscript initiatives in India and beyond. As is common knowledge, the discipline of manuscriptology encompasses the preservation, digitisation, and interpretation of handwritten texts. With the rise of digital humanities and AI-assisted palaeography, institutions worldwide have adopted diverse strategies to safeguard and disseminate manuscript heritage.

1. European Union: Interoperability and Scholarly Infrastructure

Key Initiatives:

- Europeana: Aggregates digitised cultural heritage from over 3,000 institutions across Europe, including manuscripts, books, and archival materials. (<https://www.europeana.eu>)
- Monasterium.net: Provides access to over 700,000 medieval charters with diplomatic transcriptions and palaeographic annotations. (<https://www.monasterium.net>)
- Transkribus (READ-COOP SCE): Offers AI-based Handwritten Text Recognition (HTR) for historical European scripts. (<https://readcoop.eu>)

Practices:

- IIIF for image interoperability
- TEI XML for structured transcription
- Linked Open Data for semantic search

Strengths:

- Advanced palaeographic modelling
- Collaborative networks across national libraries
- Emphasis on open access and multilingual metadata

2. United Kingdom: Conservation and Public Engagement

Key Institution:

- British Library: Hosts the Digitised Manuscripts portal, featuring collections such as the Cotton, Harley, and Royal manuscripts. (<https://www.bl.uk/manuscripts>)

Practices:

- High-resolution digitisation under conservation protocols
- Metadata encoded using MARC21, MODS, and TEI
- Public exhibitions and educational outreach

Strengths:

- Integration of scholarly commentary
- Emphasis on conservation ethics
- Outreach through crowdsourced transcription (e.g., LibCrowds)

3. United States: Technical Rigor and Distributed Access**Key Initiatives:**

- Digital Scriptorium: A consortium of US libraries offering manuscript metadata and images. (<https://digitalscriptorium.org>)
- Library of Congress Manuscript Division: Digitizes historical collections with preservation-grade imaging. (<https://www.loc.gov/rr/mss>)
- By the People: A crowdsourced transcription platform hosted by the Library of Congress. (<https://crowd.loc.gov>)

Practices:

- FADGI imaging standards aligned with ISO 19264-1
- Metadata encoded using MODS, EAD, and Dublin Core

Strengths:

- Robust technical documentation
- Distributed digitisation across institutions
- Emphasis on public domain access

4. Southeast Asia: Custodianship and Adaptive Digitisation

Key Initiatives:

- DREAMSEA: A collaboration between UIN Jakarta and the University of Hamburg, funded by the Arcadia Fund. (<https://dreamsea.co>)
- Digital Library of Lao Manuscripts (DLLM): Hosted by the National Library of Laos. (<https://laomanuscripts.net>)
- SEAlang Library: Offers linguistic and script resources for Southeast Asian languages. (<http://sealang.net>)

Practices:

- Field-based digitisation using portable imaging kits
- Metadata includes script type, language, and cultural context
- Digitisation of palm-leaf, lontar, and paper manuscripts

Strengths:

- Community involvement
- Emphasis on safeguarding endangered collections
- Integration of oral histories and ritual annotations

5. Comparative Summary

Global manuscript digitization efforts reveal diverse regional paradigms, each shaped by distinct technical standards, metadata frameworks, and cultural philosophies. European institutions emphasize scholarly infrastructure through IIF, TEI, and Linked Open Data, fostering interoperable portals grounded in TEI XML and MARC21. The UK complements this with public-facing exhibitions and conservation outreach, blending MARC21 and TEI with rich commentary to engage broader audiences. In the US, pragmatic innovation drives open-source platforms supported by FADGI, ISO 19264-1, and metadata schemas like MODS and Dublin Core. This distributed model prioritizes technical scalability and decentralized access. Southeast Asia, by contrast, champions adaptive, field-based approaches rooted in multilingual metadata and community archives, privileging custodianship and contextual integrity over standardization.

For India, these models offer both inspiration and caution. The European and UK frameworks emphasize the importance of scholarly rigor and public engagement, whereas the US model encourages scalable innovation. Yet it

is the Southeast Asian ethos—grounded in relational stewardship and local context—that resonates most deeply with India’s manuscript ecology. A hybrid Indian approach might integrate TEI and IIIF with multilingual, script-sensitive metadata, while anchoring access in community-led custodianship and regional knowledge systems. Such a model would honour both the technical and philosophical dimensions of textual transmission, integrating oral traditions and ritual contexts into metadata frameworks. By ensuring technical interoperability, it might seek to position India not merely as an adopter of global standards but as an active contributor to their evolution and enrichment. Global best practices encompass a spectrum of priorities, ranging from technical precision to cultural stewardship. By synthesizing these approaches, India can build a future-ready manuscript ecosystem that honours tradition while embracing innovation.

VIII. Way forward -- Towards a National Knowledge Movement

A time-bound, youth-led initiative is proposed to advance the preservation, decoding, dissemination, and recontextualization of India’s manuscript heritage. The objective is to catalyse the revitalization of indigenous sciences, philosophies, technologies, and ethical frameworks. This initiative is conceived not as a retrospective conservation effort, but as a forward-looking strategy for knowledge reclamation, innovation, and global leadership.

Strategic Pathways could include --

1. Youth Engagement and Employment

Establish structured fellowships and internships to train emerging scholars in manuscript decoding, palaeography, and indigenous knowledge systems. These programs shall integrate traditional expertise with digital methodologies to facilitate access to texts in Ayurveda, astronomy, statecraft, architecture, and ethics.

2. Institutional Anchoring

Mandate the creation of interdisciplinary Manuscript and Indigenous Knowledge Centres across central and state universities. These centres shall be equipped with digitization laboratories, script-learning facilities, and AI-enabled interpretation platforms to support research and pedagogy.

3. Script Integration and Linguistic Democracy

Affirm the equal epistemic value of all Indian scripts—including Brahmi,

Grantha, Sharada, Modi, Maithili, Tigalari, Tamil, Bengali, Assamese, and Devanagari. Develop transliteration matrices and AI-assisted tools to enhance script accessibility and foster linguistic inclusivity.

4. Traditional Wisdom Meets Scientific Temper

Engage traditional scholars as co-authors of manuscript interpretation protocols. Simultaneously, mobilize the expertise of scientists, technologists, and designers to build scalable infrastructure. This convergence shall ensure depth of insight and precision of delivery.

5. Global Collaboration and Civilizational Diplomacy

Position India as a global node in manuscript studies through international summits, collaborative research programs, and digitization partnerships with institutions across Southeast Asia, Europe, and North America.

Conclusion

Institutionalising manuscriptology and palaeography, along with studies in textual criticism in India, requires more than just infrastructure — it demands vision, collaboration, cultural sensitivity, and a sense of civilizational pride. By integrating policy reform, academic innovation, technological advancement, and community engagement, India can transform its manuscript heritage from a fragile legacy into a robust and living archive. Such an initiative would not only preserve the past but illuminate the future, offering generations of scholars, custodians, and citizens a deeper connection to the wisdom encoded in ink, leaf, and parchment.



Working Group IV

Theme: Digitisation Tools, Platforms, and Protocols (HTR, AI, IIF)

Coordinator

Prof. P. Jha

Head, Cultural Informatics Lab (CIL), IGNCA

Co-Coordinator

Shri Rajeev Kumar

Director, Ministry of Culture, Govt. of India

Experts

Prof. Santanu Chaudhury

Former Director, IIT Jodhpur

Prof. Rajendra Kumar Anayat

Ex-Vice Chancellor, Deenbandhu Chhotu Ram University of Science &
Technology, Murthal, Sonipat

Shri Timothy Dkhar

Deputy Director General, NIC

Prof. C. V. Jawahar

Professor, IIIT Hyderabad

Dr. Dinesh Katre

Former Senior Director & HOD, C-DAC Pune

Shri Shah Nawaz Khan

Sr. Director (IT), NIC

Platforms and Technologies for Digital Preservation and Conservation of Heritage Manuscripts under Gyan Bharatam Mission

1. Introduction

Digital technologies play a critical role in preserving and conserving India's vast manuscript heritage and its civilizational knowledge traditions for future generations. Digital preservation involves ensuring the long-term accessibility and usability of heritage materials in digital form through strategic organizational and technological measures. AI/ML tools help add value by improving the quality of digital data and providing robust transcriptions. Digital Conservation is about actively managing and maintaining the quality and integrity of digital files and their associated metadata, often involving technical repairs and enhancements. Appropriate protocol using blockchain (and complementary technologies) can support secure access and integrity management of digitized heritage manuscript collections. Both digital preservation and conservation are critical and complementary for ensuring that heritage manuscripts, once digitized, remain valuable, accessible, and authentic over time. A Trustworthy Digital Repository (TDR) recognizes the risks associated with the system (i.e., media, software, hardware, communication errors, network failures, obsolescence, operator error, internal or external attacks, organizational failures, economic failures, etc.) and mitigates those risks to make them manageable.

India boasts one of the largest collections of manuscripts in the world, estimated to exceed 10 million. These are not just documents; they are the primary sources of India's intellectual, cultural, and scientific history. They are written on fragile organic materials, such as palm leaves (Tadpatra), birch bark (Bhojpatra), handmade paper, cloth, and even wood. They encompass everything from philosophy (Vedas, Upanishads) to epics (Ramayana, Mahabharata), as well as science (astronomy, mathematics, and medicine, including Ayurveda), law, grammar, art, and literature. Digitization is a process designed to create a high-fidelity, long-lasting digital surrogate of the manuscripts. Quality of the digital data is a function of (i) material properties of the manuscripts; (ii) quality of the surface of the manuscripts; (iii) nature of binding (if bound); (iv) digitization process (camera, scanner, illumination). Resolution and colour levels used for digital representation of the manuscript data impact readability and subsequent processing of the content in the digital

space. These parameters must be set based on the nature of the manuscript, its content, and the digitization facility for optimal output.

2. OAIS Framework, ISO 14721

In the 1990s, numerous parallel efforts worldwide aimed to address the issue of digital preservation. In Washington, D.C., the Consultative Committee for Space Data Systems (CCSDS) launched a joint effort to create a reference model, known as the Blue Book. The Blue Book was published in the public domain [OAIS 02] in January 2002 and was later adopted as an ISO standard ISO 14721. In 2012, a revised and updated version of the Blue Book was published as ISO standard 14721:2012. At the core of the reference model is the concept of an OAIS (Open Archival Information System). An archive aligned with the OASIS functional model supports six high-level service elements: (i) Ingest, (ii) Archival Storage, (iii) Data Management, (iv) Preservation Planning, (v) Access, and (vi) Uninterrupted availability of data for the fulfillment of digital preservation commitment.

Recently, an updated version was released is ISO 14721:2025. The central concept in the reference model is that of an open archival information system. Some important features of the OAIS Reference Model are placed below: -

- The central concept in the reference model is that of an open archival information system. An OAIS-type archive must meet a set of six minimum responsibilities related to the ingest, preservation, and dissemination of archived materials.
- An OAIS-type archive operates in an environment populated by three types of entities: Management, Producer, and Consumer. A special class of consumer is called the Designated Community: the subset of consumers expected to independently understand the archived information in the form in which it is preserved and made available by the OAIS. An OAIS-type archive's external environment could also include interaction with other OAIS archives.
- The reference model identifies and describes the core set of mechanisms with which an OAIS-type archive meets its primary mission of preserving information over the long term and making it available to the Designated Community. These mechanisms are summarized by the OAIS functional model, which defines six high-level services, or functional entities, that collectively define the OAIS's preservation and access operations: Ingest,

Archival Storage, Data Management, Preservation Planning, Access, and Administration. Operating alongside these six functional entities are Common Services, which consist of basic computing and networking resources. An OAIS-type archive will implement each of the six functional entities, along with Common Services, in the course of building a complete archival system.

- The reference model provides a high-level description of the information objects managed by an OAIS-type archive. The OAIS information model is built around the concept of an information package, which consists of the object that is the focus of preservation, along with metadata necessary to support its long-term preservation, access, and understandability, bound into a single logical package. There are three important variants of the information package concept: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP).
- The OAIS reference model includes a discussion of different classes of interoperability across OAIS-type archives: independent archives, cooperating archives, and federated archives. The reference model also notes that archives can interoperate through shared functional areas.
- A number of initiatives have used the OAIS reference model as a conceptual foundation and starting point for more focused work in digital preservation. Areas of application include, but are not limited to, ‘OAIS-compliant’ repository architectures and systems; repository self-assessment and certification; metadata requirements for digital preservation; methods and protocols for encoding and exchanging archived information; and other OAIS-related standards.
- Because the reference model is a conceptual framework rather than a blueprint for concrete implementation, the meaning of ‘OAIS-compliant’ is necessarily vague and open to interpretation. A key element in the design of OAIS is its flexibility and level of abstraction: it makes no assumptions about how the concepts and models in OAIS are to be implemented, and imposes no requirements concerning the technologies used to support the implementations. Despite the attendant ambiguity, the notion of OAIS conformance has been beneficial, to the extent that it helps consolidate

understanding of the fundamental requirements for securing the long-term persistence of digital materials – a necessary condition for building well-understood, interoperable, and ultimately, trusted digital preservation systems.

- Perhaps the most important achievement of the OAIS reference model to date is that it has become almost universally accepted as the lingua franca of digital preservation, shaping and sustaining conversations about digital preservation across disparate domains, and supplying a general mapping of the landscape that stewards of our digital heritage must navigate in order to secure the long-term availability of digital materials. Alignment with concepts defined in OAIS helps orient a technical implementation, draft standard, or other activity within the broader repository context that the OAIS reference model defines, making it part of a cohesive 'big picture'. It seems reasonable to conclude that OAIS has become a foundational resource for understanding digital preservation, a language for discussing digital preservation issues, and a starting point for implementing digital preservation solutions.
- It is possible to identify a few limitations associated with OAIS's impact. Very few of its concepts have been directly and formally operationalized as standards in their own right. A design, a protocol, or even a standard can self-declare itself OAIS-conformant (but without an explicit accounting of how conformance is actually manifested). Initiatives can use OAIS concepts as a means of labelling or describing various components within their structure (but these concepts can be used quite superficially, more as an expositional shorthand rather than a detailed mapping); OAIS can be cited as a foundation or starting point for a particular initiative, or alternatively the initiative can declare itself informed by OAIS (but without necessarily any elaboration on how this was so). It is useful to remember that an OAIS-type archive is still one built primarily on OAIS concepts, not an OAIS suite of standards. The digital preservation community would benefit from a careful assessment of where more precise and authoritative definitions of OAIS concepts and relationships would accelerate progress in achieving robust, widely applicable, and interoperable digital preservation solutions.

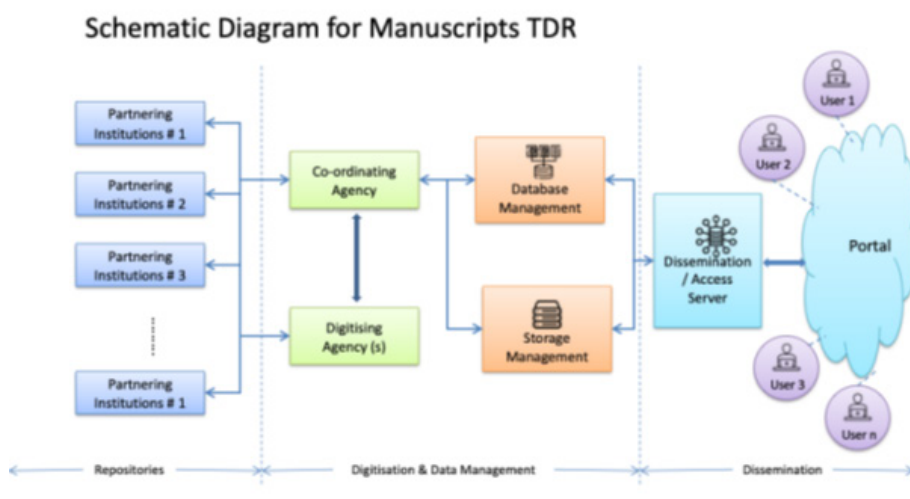
3. Trusted Digital Repositories

Trusted Digital Repositories (TDRs) are evaluated and certified by third parties to provide a high level of assurance to users that the archive receives and maintains all the digital documents submitted to it and that if accessed in the future, it will disseminate the same digital document (Ambacher 99). The first Trusted/Trusted DDR standards were published in February 2007 under the title, 'Evaluating Audit & Certification Criteria Towards an International Auditing & Certification Process Using Checklist for Audit & Certification'. The revised and updated version of the standards for TDRs was published in 2012 under ISO standard 16363:2012. The following three repositories worldwide have been certified so far.

- National Cultural Audiovisual Archives (NCAA) project of the Ministry of Culture, Govt. of India, implemented by Indira Gandhi National Centre for the Arts with technical support from C-DAC Pune, has received ISO 16363:2012 certification to become the world's first TDR in November 2017.
- The US Government Publishing Office Govinfo repository has been awarded global ISO 16363 certification by PTAB. The U.S. Government Publishing Office (GPO) makes history by becoming the first organization in the United States and the second organization worldwide to achieve the highest global standard of excellence for digital repositories. The Primary Trustworthy Digital Repository Authorization Body Ltd. awarded GPO ISO 16363:2012 for govinfo, the one-stop site to authentic, published government information. GPO achieved certification by meeting the official criteria for trustworthy repositories, as defined by experts in the field.
- ETERNAL RDC-Arq Digital Repository of Brazil has received ISO 16363:2012 certification to become the world's third TDR in February 2024. The certification for the repositories is valid for three years, and it must undergo a fresh audit for recertification. With technological advancements, these standards are also being revised, and the repositories need to undergo recertification. The last revised and updated version of the standards for TDRs was published in 2025 under ISO standard 16363:2025.
- Trustworthiness refers to a system's ability to perform according to its stated goals and requirements, i.e., to fulfill its promises as intended. When assessing the trustworthiness of a repository, various factors must be taken into consideration, such as Organizational structure (governance,

policy, and process), financial stability and sustainability, licenses, and liabilities of a trustworthy inheritor of data. Additionally, to ensure integrity, authenticity, confidentiality, and availability of data, its digital object management practices, technological infrastructure, and data security measures must be adequate to fulfill the commitment of digital preservation. A TDR, on the other hand, recognizes the risks associated with the system (i.e., media, software, hardware, communication errors, network failure, obsolescence, operator error, internal or external attack, organizational failure, economic failure, etc.) and turns those risks into manageable risks.

- Digital Preservation is a system for managing digital information over time. It is defined as the long-term, error-free storage of digital information for retrieval and interpretation. The majority of the information (Textual, Visual, Audiovisual, etc.) generated today is in digital form. The unique characteristic of the digital makes it easy to create and keep content up-to-date, but at the same time, it presents several difficulties in preserving this content. It's no secret that digital preservation takes more time and effort than preserving other media. We require constant investment of time, effort, and money to keep up with the rapid pace of technological and organizational development, which is often considered the biggest obstacle to digital preservation. In fact, while we can still access our written heritage from thousands of years ago, digital information created just 10 years ago is at risk of disappearing. Digital preservation poses challenges of a completely different nature from those associated with preserving data in traditional formats.



The biggest challenge digital preservation faces are that the media on which digital contents are stored are more vulnerable to deterioration and catastrophic loss than most analog media. The recording media for digital data deteriorate at a much faster pace, and once deterioration begins, data loss typically occurs in most cases. Digital obsolescence is another challenge, perhaps a more serious and important one, as it directly relates to the problem of long-term access. This challenge is exacerbated by the lack of established standards, protocols, and proven methods for preserving digital information. There are several strategies, such as refreshing, Migration, Replication, Emulation, Metadata Integration, and Trustworthiness of digital objects, that can be used to actively combat the loss of digital information.

Manuscripts in analog form have been preserved for many centuries. Gyan Bharatam Mission proposes to digitize them in a time-bound manner for the benefit of scholars and promotion of the traditional Indian Knowledge System. Today, technology offers various features, such as searching and HTR, which can help scholars in exploring and analyzing the content enshrined in manuscripts. At the same time, the long-term preservation of digitized data is also a crucial consideration when planning for digitization, based on international standards.

A TDR is a well-established process for the long-term preservation of digital data. However, there is always a finite possibility of organizational and/or financial failure of a TDR. The process ensures data recovery in the event of any disaster. Additionally, there is always the possibility of reconfiguring another TDR as its trusted inheritor, which, in the event of its failure, would preserve its data. Necessary infrastructure is required for institutions to establish digital repositories.

4. Digitization of Manuscripts and enhancing their readability in the Digital Domain

4.1 Digitisation Process

Image Capture: This is the core technical step that determines the quality of the data. This process may address the following key concerns

- **Equipment:** High-end overhead scanners are preferred over flatbed scanners. They are non-contact, which prevents damage to the fragile manuscripts and their bindings. These criteria help ensure that each tool is used in the most efficient and effective way, preserving the quality of the material while minimizing potential damage.
- **Resolution:** Images are to be captured at very high resolutions (e.g., 600

DPI or higher using, for example, 16- or 32-megapixel overhead capture devices like CZUR Document Scanner – ET 18 Pro) to capture the finest details of the script and the texture of the material for old and rare manuscripts.

- **Lighting:** Controlled, cool, UV-filtered LED lighting is used to prevent damage and ensure even illumination.
- **Colour Accuracy:** A colour calibration chart (like a Color Checker) to be included in a reference shot to ensure the digital colours are true to the original.
- **File Format:** The master copies are saved in a lossless format like TIFF (Tagged Image File Format) to preserve all the captured data. Access copies for web viewing are created in formats like JPEG or PNG.

It is desirable to scan in Color or Grayscale and not scan directly to black and white (1-bit). Scanning in grayscale or color captures all the tonal information (stains, faded text), providing more data to work with during restoration. Desirable to use a Black Background behind the page/leaf being scanned. This can drastically reduce bleed-through from the other side.

For rare and ancient manuscripts, we utilize hyperspectral and multispectral imaging to uncover hidden yet critical information. about our heritage of manuscripts and associated knowledge. The following table highlights the utility of these – MSI (Multi-spectral) and HSI (hyper-spectral) imaging modalities:

| Application | Description | Technology Best Suited |
|---|--|--|
| Revealing Faded or Damaged Text | Faded inks, which are barely visible to the eye, often have a strong spectral signature in the near-infrared. This allows researchers to "see" the text or digitize it as if it were freshly written. | Both MSI and HSI work well. MSI is often sufficient. |
| Reading Palimpsests | A palimpsest is a manuscript where the original text was scraped off and a new text written over it. The spectral signatures of the residual, underlying ink, and the new ink are different, allowing computers to digitally separate the two layers of text. | HSI is superior due to its spectral resolution, which can distinguish between two very similar inks. |
| Material Characterization (Inks & Pigments) | Different inks and pigments have unique spectral signatures. HSI can identify them non-destructively, helping to date a manuscript, identify an artist's palette, or detect forgeries. | HSI is the gold standard for this. |
| Assessing Condition & Conservation | These techniques can identify areas of mold, foxing, water damage, or previous conservation efforts that are not visible to the naked eye. This helps in planning conservation treatments. | Both can be used, but HSI provides more detailed chemical information. |
| Mapping and Visualization | By processing the data cube, researchers can create "false color" images that assign a specific color to a specific material. For example, all text written in iron gall ink could be colored red, and all text in carbon black could be colored blue, instantly showing where different scribes worked. | HSI provides the data necessary for this sophisticated analysis. |

Table: 01

We need a major initiative to utilize advanced imaging techniques for significant manuscript collections, preserving our manuscripts in the digital domain and making them available to researchers for analysis. This will enable these snippets of our knowledge system to be globally accessible. Effectively, for true digital conservation of manuscripts, it is extremely important to capture data so that a Digital Twin of the manuscript can be created. A digital twin of a manuscript will enable all subsequent research for a deeper understanding of the manuscript (not just the text but its historical and sociological context) using the digital twin and not accessing the physical manuscript – ensuring

long-term preservation of the manuscript.

4.2 Better Readability of Manuscripts Using Image Enhancement

Various physical factors can affect the degradation of document quality over time, including the material quality of the writing or print surface, decolouration over time, paper quality, and environmental factors such as mold and infestation. Once a high-quality digital image is created, computational techniques can be used to make the text significantly clearer than it might appear to the naked eye on the original manuscript.

4.2.1 Common Problems in Scanned Images of manuscripts

- i. Skew / Rotation: The document was not perfectly aligned on the scanner, resulting in a tilted image.
- ii. Noise: Random speckles (“salt-and-pepper” noise) or grainy textures, often from old paper or low-quality scanning.
- iii. Low Contrast & Faded Text: The text is faint and blends into the background, common in old manuscripts.
- iv. Stains, Foxing, and Discoloration: Yellowing paper, water stains, coffee rings, or reddish-brown spots (foxing) from age and humidity.
- v. Bleed-through / Show-through: Text or images from the reverse side of the page are visible.
- vi. Uneven Illumination: The image is darker at the edges and brighter in the center, often from using a camera instead of a flatbed scanner.
- vii. Fold Lines and Creases: Physical folds in the paper create dark lines and shadows in the scan.
- viii. Low Resolution or Blurriness: If the scan was performed at a low DPI, it may result in pixelated or blurry text.

4.2.2 Restoration Process

Restoration workflows are designed to minimize degradations in the manuscript images. The following steps are required:

1. Cleaning and Repair: This is the process for removing noise, including stains and blemishes.
2. Color and Tone Correction: This stage eliminates noise and highlights the regions of interest – typically, textual regions. Removes colour casts caused by aging paper or inconsistent lighting. This can make the background more uniform, allowing the text to be more visible.
3. Contrast and Brightness Adjustment: Faded ink can be made to stand out against the background.

4. Sharpening: Filters can enhance the edges of the script, making the characters appear crisper and easier to read.
5. Deskewing (Straightening): Rotate the image so that the lines of text are perfectly horizontal. Most image editing software has an automatic or manual straightening tool. This is critical for accurate OCR.
6. Binarization (Thresholding): This is the process of converting a grayscale image into a pure black-and-white image (1-bit). This is excellent for creating clean, highly readable text.

4.2.3 State-of-the-art models for manuscript restoration

Advanced AI/ML models are available for enhancing the readability of scanned manuscripts. However, the challenge of using ML for the processing of manuscripts is the availability of sufficient training data. Therefore, we propose to use the Generate, Transform, and Clean (GTC) strategy to address these critical tasks. The GTC strategy encompasses a systematic process aimed at generating augmented data and performing the noise cleaning or binarization task, offering a comprehensive solution to these challenges. In the “Generate” stage, we use generative AI techniques to generate realistic background datasets. The “Transform” stage is designed to merge the cleaned text with these generated backgrounds, thereby producing diverse datasets. By creating such a varied and enriched dataset, this stage significantly increases the learning capacity of our model, accommodating a wide range of manuscript conditions and variations. These datasets are then prepared for the final machine learning based “Clean” stage. AI models are far superior to traditional filters at removing noise without blurring the actual text, as they have been trained on what characters should look like.

[Nimol Thuon, Jun Du, ·Zhenrong Zhang, Jiefeng Ma, ·Pengfei Hu, Generate, transform, and clean: the role of GANs and transformers in Palm leaf manuscript generation and enhancement, IJDAR, (2024) 27:415–432]

4.2.3.1 State of the Art Binarization:

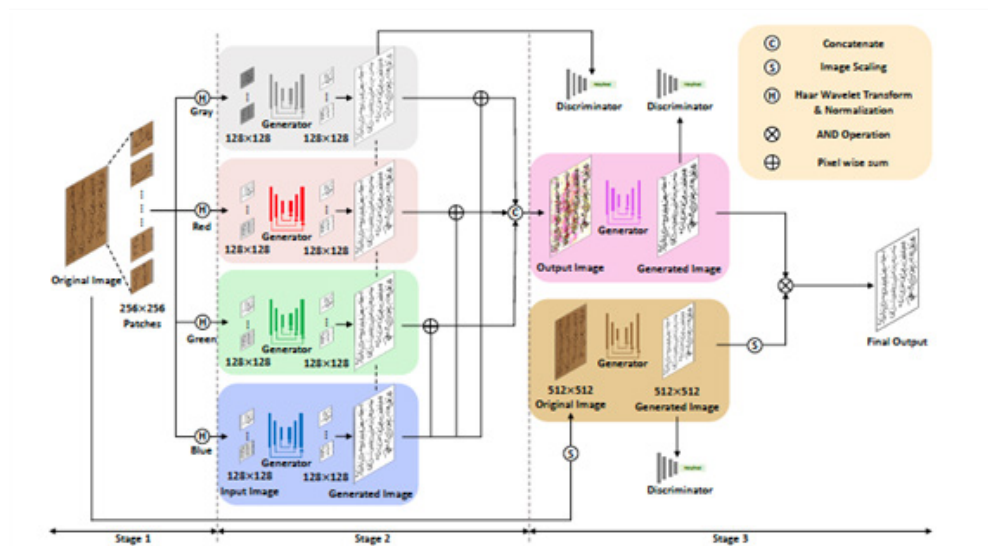
Traditional methods of converting an image to black and white (binarization) struggle with uneven backgrounds. ML models can be trained to intelligently distinguish between text and background, even on heavily stained or damaged pages. This is a critical pre-processing step for Optical Character Recognition (OCR).

Global thresholding (Otsu’s method) and Local/Adaptive thresholding methods are traditionally used. However, global methods fail in cases of uneven illumination, shadows, stains, or faded areas. While better, local methods are

highly sensitive to their parameters (like the size of the local window). They can still produce a lot of noise, broken character strokes, or thick, merged characters, especially on heavily degraded historical documents.

AI, specifically Deep Learning, has revolutionized document binarization. Instead of relying on hand-crafted rules about pixel brightness, an AI model learns what constitutes text and what constitutes background from examples. This approach treats binarization as an image segmentation problem: The model's job is to classify every single pixel in the input image as either "foreground" or "background". Generative adversarial networks (GANs) can generate images where shadows and noise are effectively removed, allowing for the extraction of text information. A GAN-based three-stage network that incorporates a discrete wavelet transform provides state-of-the-art results with reduced training and inference time.

[Efficient GANs for Document Image Binarization Based on DWT and Normalization. RY Ju, KS Wong, JS Chiang. arXiv preprint arXiv:2407.04231, 2024, GitHub at https://github.com/RuiyangJu/Efficient_Document_Image_Binarization.]



The novel three-stage network architecture of the proposed method for document image binarization: image processing (Stage 1), document image enhancement (Stage 2), and document image binarization (Stage 3).

The model is shown an input image (e.g., a photo of a noisy image). It produces its best guess for the binary output. This output is then compared to the ground truth binary image. The difference (the “error” or “loss”) is used to adjust the model’s internal parameters via backpropagation. This process is repeated millions of times until the model becomes extremely accurate. Inference: Once trained, the model can take a new, unseen document image and, in a single forward pass, produce a high-quality binary output.

There are open-source OCR engines that feature a powerful binarization module, which can be used independently. Those use U-Net-like architectures.

- **ocropy/ocropus:** An older but still relevant OCR toolkit that has advanced binarization capabilities.
- **OpenCV:** While it primarily contains traditional methods (`cv2.adaptiveThreshold`), you can use its deep learning module (`cv2.dnn`) to run a pre-trained U-Net model saved in ONNX format.

4.2.3.2 Denoising: State of the Art

Unlike noise in natural photos (e.g., low-light grain), document noise is highly varied and can be structural in nature. Simply blurring an image would destroy the text.

Common Types of Document Noise:

- **Salt-and-Pepper:** Random black and white pixels.
- **Gaussian Noise:** Random variations in intensity, often from scanner sensors.
- **Ink Bleed-through / Show-through:** Text from the reverse side of the page is visible.
- **Stains and Blotches:** Coffee stains, watermarks, mold.
- **Creases, Folds, and Shadows:** Structural artifacts that create dark lines or gradients.
- **Low Contrast / Fading:** Old documents where the ink has faded against the background.
- **Background Texture:** The texture of the paper itself (e.g., papyrus, recycled paper).

Traditional methods like Median Filters, Gaussian Blur, or Thresholding (Binarization) struggle with this complexity:

- * They are not context-aware.
- * They often blur sharp edges, which is disastrous for text and reduces Optical Character Recognition (OCR) accuracy.

- * They require manual parameter tuning for each different type of noise.
- * They cannot effectively handle complex, non-uniform noise, such as bleed-through or shadows.

AI models, particularly deep neural networks, learn the underlying structure of a “clean” document. They learn what text is expected to look like, what a straight line is, and what a uniform background should be. By understanding the “ideal,” they can effectively identify and remove anything that deviates from it—the noise. Some of the approaches being used are:

a) Convolutional Autoencoders (CAE)

This is the basic architecture for denoising.

- **Encoder:** A series of convolutional layers that compress the noisy input image into a compact, low-dimensional representation called a latent vector. In this process, the network is forced to learn the most essential features (the text and layout) while discarding the non-essential information (the noise).
- **Decoder:** A series of deconvolutional (or upsampling) layers that reconstruct the image from the latent vector. Because the noise was discarded, the decoder reconstructs a clean version of the document.

[Kulkarni, et al.2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Pune, India. Apr 7-9, 2023)

b) Generative Adversarial Networks (GANs)

GANs offer state-of-the-art performance by using a competitive, two-player system.

- **The Generator:** This network’s job is to create clean document images. It takes a noisy image as input and attempts to produce a denoised output that appears realistic. Often, the generator itself has a U-Net-like structure.
- **The Discriminator:** This network is trained to distinguish between real clean documents and the fake denoised images produced by the generator.
- **The Training Process:** The generator and discriminator are trained in a continuous loop. The generator gets better at fooling the discriminator, and the discriminator gets better at catching fakes. This adversarial process pushes the generator to produce extremely realistic and artifact-free clean documents. GANs are particularly good at “in-painting” areas destroyed by large stains or creases.

c) Mixture of Experts

The noise cleaning problem can be formulated as follows: given two unpaired

sets of documents, one set consisting of noisy documents (X) and the other a collection of clean documents (Y), and knowing that these two sets are unpaired, can we transform one set to have the style of the other? This problem can be formulated as an unsupervised image-to-image translation problem. Recently, several proposals have been made to address this problem, with one of the most prominent being based on cycle-consistent GANs (also known as cycle-GANs). In fact, in this solution, we remove/restore noise/ degradation from these documents while preserving their text contents. The noisy documents X may be corrupted by several different types of noise. Integration of a deep mixture of experts for different types of noise into the cycle-GAN provides a robust solution. A cycle-GAN consists of two generators: forward (H) and backward (F) generators, and two adversarial discriminators, DY and DX.

The generators transform the data from one domain to another, i.e., $H: X \rightarrow Y$ and $F: Y \rightarrow X$. The adversarial discriminators aim to differentiate between the outputs of generators and the real data, i.e., DY aims to discriminate between $H(x)$ and y , whereas DX tries to distinguish between $F(y)$ and x . The objective function in cycle-GAN is based on two losses: the GAN loss, which transforms the image style from one domain to another, and the cycle consistency loss, which preserves the image content. To design a single model that can tackle different noise types, a model of a mixture of experts can be integrated into the cycle-GAN. A simple approach is to combine individual trained cycle-GANs (one for each noise type) with an ensemble learning approach on top.

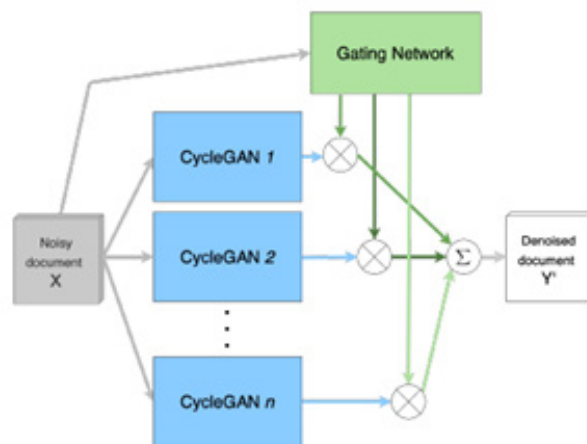


Chart: 03

[Mehrdad J Gangeh, Marcin Plata, Hamid R Motahari Nezhad, Nigel P Duffy: End-to-End Unsupervised Document Image Blind Denoising - IEEE/CVF International Conference on Computer Vision (ICCV), October 2021.]

d) Transformer-Based Approaches

Vision Transformers, utilizing an encoder-decoder architecture, have been employed to enhance and binarize manuscript images in an end-to-end fashion. The encoder operates directly on the pixel patches with their positional information, while the decoder reconstructs a clean image from the encoded patches. The input image is split into patches, which are linearly embedded, and the positional information is added to them. The resulting sequence of vectors is fed to a standard Transformer encoder to obtain the latent representations. These representations are fed to another Transformer, representing the decoder, to obtain the decoded vector, which is then linearly projected into vectors of pixels representing the output image patches.

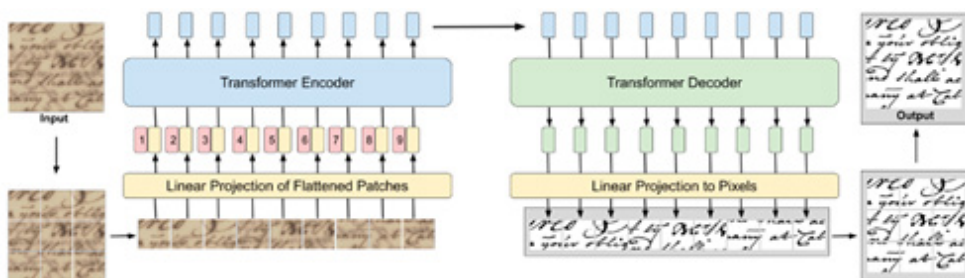


Chart: 04

[Mohamed Ali Souibgui, et. al., DocEnTr: An End-to-End Document Image Enhancement Transformer, 26th International Conference on Pattern Recognition (ICPR) August 21-25, 2022, Montréal, Québec, Canada,

4.2.4 Diffusion Models: Advanced AI Techniques

Advancements in AI techniques, such as diffusion models, provide a powerful method for enhancing the readability of manuscripts. We developed a novel Latent Alignment Head that aligns semantic features across noisy and clean domains within the diffusion process. By enforcing consistency not just at the pixel level but also in the latent feature space, the model better preserves fine-grained textual and structural details while effectively removing noise. The

bidirectional nature of the cyclic framework ensures robust domain translation and latent alignment, thereby improving the semantic fidelity of restored outputs. Evaluated on real-world datasets of degraded Indian scripts, our method outperforms traditional and diffusion-based baselines, demonstrating its effectiveness in restoring legibility and maintaining historical integrity. Sample result for a noisy manuscript preserved at IGNCA is presented below.

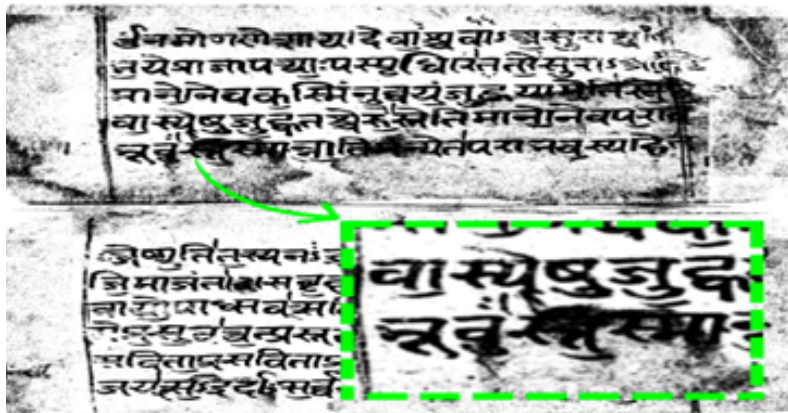


Figure 2: Noisy Image

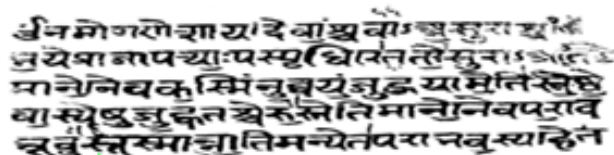


Figure 01: Our Results

4.2.4 Virtual Unfolding

For scrolls or warped/damaged pages that cannot be physically flattened, 3D scanning (using CT scans or structured light) can create a digital 3D model. Algorithms can then “virtually unroll” or “flatten” the manuscript, making the text readable.

5.0 Recognition and Retrieval

Today’s AI algorithms can be adapted to design effective algorithms for pre-processing and segmentation. More details on pre-processing is available in another section of this article. However, a major bottleneck is the lack of annotated ground-truth data for training AI models.

This is true for both pre-processing and segmentation. For pre-processing,

often, the truth is not even available. For segmentation, this demands manual annotation, which is extremely labour-intensive and subjective, with even experts disagreeing on line or word boundaries. Errors in pre-processing and segmentation directly propagate downstream, severely affecting recognition, indexing, and transcription, making these steps critical yet non-trivial for any computational pipeline aimed at preserving and making accessible the vast corpus of Indian manuscripts.

5.1 Recognition of Handwritten Text

Handwritten OCR has become increasingly practical with the emergence of commercial APIs, solutions, and applications. However, its widespread use remains predominantly restricted to English and other Latin scripts, with only limited progress in certain Asian languages that use scripts. A major bottleneck is the scarcity of large, high-quality datasets and the limited sharing of resources, which prevents the extension of OCR capabilities to new languages, particularly Indian languages. Even for English, the number of publicly accessible handwritten datasets is small, leaving academic research trailing behind industry-driven solutions. While comprehensive recognition and understanding of handwritten documents remain an open challenge, modern document understanding systems have evolved to focus on user-centric tasks such as Question Answering (QA). This paradigm shift presents a timely opportunity to explore information extraction from handwritten documents, where recognition accuracy and structural layout interpretation are equally critical.

Handwritten Text Recognition (HTR) in a low-resource scenario (where labelled data is limited) remains a difficult challenge. This is especially true for historically encrypted manuscripts, or ciphers, which were often used in military, diplomatic, or private communication. To conceal their contents, authors devised unique writing systems and alphabets, leading to high variability in style and structure. While approaches such as few-shot learning and self-supervised learning offer potential solutions to data scarcity, current HTR models still fall short of achieving reliable performance in this domain. There is a need to develop robust HTR models for Indian manuscripts, which are available in various languages, scripts, and writing styles. Two parallel efforts need to be explored: (a) creating annotated data with minimal human (expert) effort. (b) Use of a large volume of unlabelled data using self-

supervised and semi-supervised learning.

Handwritten text recognition introduces many unique challenges that make it more demanding than printed text recognition. First, handwriting exhibits significant style variability, i.e., writers differ widely in letter shapes, spacing, and slant, which makes designing robust recognition algorithms highly complex. Second, there is content variability, since handwritten material may range from formal records and academic texts to casual notes, each with its own conventions and ambiguities. Third, temporal changes add another layer of complexity, as an individual's handwriting may evolve over time, requiring recognition models to adapt dynamically. Fourth, the evolution of language and script over time makes reading challenging even for humans. These challenges, while formidable, continue to motivate researchers, underscoring the interdisciplinary depth and dynamism of handwritten OCR. At its core, OCR bridges the visual and machine-readable domains, with handwritten recognition remaining one of the most intricate frontiers of this pursuit.

Despite remarkable progress in languages such as English, Chinese, Arabic, and Japanese, a considerable gap persists for many global languages, particularly those spoken in India. Only a few of the 22 scheduled Indian languages have received focused attention in OCR research, and even then, efforts are typically limited to printed text. The scarcity of progress in handwritten recognition for Indic scripts places these languages at risk of exclusion from future technological advances. The need for research is particularly urgent for widely spoken languages such as Hindi, Bengali, and Telugu. In manuscripts, there are many more language scripts derived from Sanskrit and ancient Indic scripts that present unique challenges, which go far beyond those of the Latin alphabet. A defining feature is the prevalence of conjunct characters, where two or more characters combine into complex ligatures, dramatically increasing the difficulty of recognition. Moreover, while English consists of only 52 upper- and lowercase characters, most Indic scripts contain well over 100 basic Unicode characters, excluding modifiers, diacritics, and conjuncts. This richness in character sets, coupled with script-specific writing rules, makes the task of building accurate OCR systems for Indian languages significantly more demanding and necessitates specialized efforts.

Today, we have AI algorithms that can learn from a small number of examples and generalize to many practical situations. We need to explore such data augmentation schemes. Beyond recognizing text, there is a need to develop

a robust word/line segmentation module that can detect them in the real world. A research effort in this direction needs to focus on (i) image pre-processing, (ii) segmentation, (iii) reading order, (iv) recognition of text, and (v) enhancing accuracy with domain knowledge.

5.2. Retrieval of Handwritten Manuscripts

In recent years, keyword spotting (KWS) has emerged as a highly effective strategy for indexing large collections of handwritten texts and enabling large-scale information retrieval. Unlike full transcription, KWS focuses on detecting and retrieving specific query words directly from document images. This makes it significantly more accurate and practical than OCR-based methods, particularly when the goal is to locate keywords rather than generate complete transcriptions. The process involves matching words in document images to a given query word and ranking them based on similarity to the query representation. In KWS, users provide a textual query (single or multi-word) and are accurately located in the manuscript collection. In many cases, indexing is carried out for computational efficiency. While the query is a text, the retrieved is an image patch and a manuscript where the query is grounded. For many users, this grounding is crucial, as they can now access the relevant manuscript and appreciate its content.

KWS approaches are typically categorized along three dimensions. First, based on the query type, query-by-example (QBE) methods use an image of a word as input, while query-by-string (QBS) methods take a textual string as input. Second, based on methodology: segmentation-based approaches require words to be segmented beforehand and match the query with these segmented units, while segmentation-free approaches operate directly on the full document image without prior segmentation. Third, based on learning strategy: learning-free methods rely on hand-crafted features and direct similarity measures between query and target words, whereas learning-based methods leverage machine learning models trained on large collections of word images to enhance retrieval performance. Over the years, KWS has attracted considerable research attention in both printed documents and manuscripts. The image patches can be represented using classical image representations or compact, deep, learned representations. The challenge is in efficiently searching over the collection (similar to the sliding window schemes in object detection). For the large collection of Indian manuscripts, we need to provide retrieval/access methods based on (a) metadata, (b) keywords, and

(c) rich information content. The third is often carried out today with the help of powerful language models, which are still in their infancy for Indian languages. We briefly describe this direction in the next section.

5.2.3 Semantic Access to Manuscripts

Semantic access to the manuscripts requires strong language models and external knowledge. Unlike traditional information retrieval (IR) systems that primarily focus on retrieving relevant documents, LLM-based question answering (Q&A) systems aim to provide concise and accurate answers directly, often synthesizing information across multiple sources. A key architectural paradigm in this space is Retrieval-Augmented Generation (RAG), which integrates the generative capabilities of LLMs with retrieval techniques. In such systems, relevant documents or passages are first retrieved from large text corpora using vector-based search, after which the LLM generates precise, context-aware answers grounded in the retrieved content. This hybrid approach mitigates the limitations of standalone language models by anchoring their outputs in factual evidence, thereby improving reliability. Another crucial design dimension is prompt engineering, which investigates strategies for crafting effective prompts to guide model behavior, elicit desired responses, and reduce undesirable outputs. Despite these advancements, practical challenges remain. Chief among them is the problem of hallucination, where LLMs generate factually incorrect information—often with unwarranted confidence.

For effective access to the manuscript collection, we need to envisage a practical system that uses the power of modern AI and language models. This is a direction for research that is achievable in the coming years, given the current pace of development in the area. The proposed system transcends traditional keyword-based search by incorporating advanced semantic search capabilities. It should understand the context, intent, and meaning behind user queries, rather than merely matching words. For instance, a search for “educational model” will surface relevant documents even if the exact phrase does not appear, by recognizing related concepts such as “schooling,” “university,” or “teaching,” etc. This enables a holistic understanding of queries, capturing synonyms, paraphrases, and relationships between topics. To ensure linguistic inclusivity, the platform will support input and output in all 22 official Indian languages listed in Schedule VIII of the Constitution. Users can pose questions in their native language and receive meaningful, synthesized responses in the same language, regardless of the original

language of the debate content. This cross-lingual capability dismantles language barriers, democratizing access to legislative information for a truly pan-Indian audience. At the heart of the system is a citation-first philosophy: every generated response will be explicitly grounded in actual passages from the Vidhan Sabha debates. Instead of vague summaries, answers will include verbatim excerpts with clear Session, Page, and Paragraph identifiers, allowing users to trace information directly to its source. This approach fosters transparency, reliability, and academic integrity, making the platform invaluable for journalists, researchers, policy analysts, and other professionals. In situations where sufficient evidence is unavailable, the system will gracefully abstain, clearly indicating that it could not find a confident, supported answer, rather than providing speculative information.

6.0 Conclusions

The success of the Gyan Bharatam project is critically linked to the development and deployment of appropriate technologies for making our heritage manuscripts widely accessible to the next generation in the digital space. This will require a mission-mode initiative for technology development and appropriate translation to create a distributed digital library of our heritage knowledge. Ensuring trustworthy digital repositories and readable, usable content across linguistic boundaries will require significant technological development efforts, for which Indian researchers and technologists can make a substantial contribution. Gyan Bhramam mission has provided an opportunity to revisit our past and understand our heritage using technology-driven tools and methodologies. The mission must adopt standards and enable the development of appropriate technology to reach the goal. The challenges and scale of this mission require innovations, not just simple project implementation.



Working Group V
Theme: Conservation and Restoration of Manuscripts

Coordinator
Dr. Achal Pandya
Prof. & HoD (Conservation), IGNCA

Co-Coordinator
Shri Dalpat Singh
Director, Ministry of Culture, Govt. of India

Experts
Prof. Sukhdev Singh
Vice Chairman, INTACH

Dr. P. Perumal
Former Conservator & Librarian, Sarasvati Mahal Library, Thanjavur

Dr. Sanjay Dhar
Expert Conservator, IGNCA

Prof. K. K. Jain
Former Professor, National Museum Institute (NMI)

Ms. Ritu Jain
Chief Conservator, Parliament Museum and Archives

Shri Nilabh Sinha
Principal Director, INTACH Conservation Institute (ICI), Arts and Material
Heritage

Dr. Tej Singh
Former Director, NRLC, Lucknow

Safeguarding India's Written Heritage: Preventive and Remedial Conservation of Manuscripts and Copper Plate Inscriptions

Introduction

India boasts one of the richest traditions of written heritage in the world, encompassing manuscripts, rare books, archival documents, and printed materials that date back centuries. These materials are not merely physical objects; they are carriers of knowledge, wisdom, and cultural identity, reflecting the intellectual, spiritual, and artistic achievements of generations. From palm-leaf manuscripts containing ancient philosophical treatises to early printed books chronicling the colonial era, the diversity of formats, materials, and subjects is vast.

However, the preservation of this heritage faces a pressing challenge: while the volume of manuscripts and books is enormous, the number of trained conservators in India is alarmingly small. This imbalance creates an urgent need for a conservation strategy that is both practical and sustainable. Preventive conservation emerges as the most viable and impactful approach in such a situation.

Preventive Conservation

ICOM-CC Definition of Preventive Conservation) Preventive conservation refers to the set of measures and actions aimed at avoiding or minimising future deterioration or loss of cultural heritage. Unlike remedial conservation, which involves direct physical intervention to stabilise or restore an object, preventive conservation focuses on controlling the environment, handling, storage, and usage conditions to ensure the long-term safety of materials.

In the context of books and manuscripts, preventive conservation does not require advanced laboratory facilities or highly specialised equipment. With basic training, custodians, librarians, archivists, and even volunteers can effectively implement it. This is particularly important in India, where many collections are housed in small libraries, temples, monasteries, private homes, and institutions that lack professional conservators.

Indigenous materials can be identified using simple testing kits. One such testing methodology is the Oddy Test.

Why Preventive Conservation is important for India

- 1. Sheer Volume of Collections** - India is estimated to have millions of manuscripts in Sanskrit, Persian, Arabic, Pali, Tamil, and other languages, scattered across public and private repositories. Many collections remain uncatalogued and are vulnerable to neglect. It is logistically and financially impossible to provide remedial conservation for every damaged item, making preventive care a more feasible large-scale strategy.
- 2. Shortage of Trained Conservators** - Only a limited number of trained paper and book conservators work in India, with the majority concentrated mainly in major urban centers or national institutions. Smaller institutions and local repositories often rely on untrained staff or well-meaning volunteers who may unintentionally cause harm.
- 3. Risks of Untrained Conservation Work** - Conservation carried out without proper training—whether remedial or preventive—can lead to irreversible damage. Examples include the use of adhesive tapes that stain and weaken paper, oil-based polishes that attract dust, and inappropriate cleaning methods that abrade surfaces or fade inks. We have also seen in the past the rampant use of cellulose acetate lamination, making many manuscripts beyond repair. In some cases, the damage caused exceeds natural deterioration.
- 4. Climatic and Environmental Factors** - India's tropical and subtropical climates present specific preservation challenges, including high humidity, fluctuating temperatures, monsoon dampness, insect infestations, and dust. Preventive measures, if applied systematically, can mitigate these threats without costly interventions.

Preventive conservation in India must combine global best practices with local traditional wisdom. With basic training, custodians can adopt the following measures:

- 1. Good Housekeeping Practices** - Clean storage spaces are essential. Dust, dirt, and organic debris attract insects and encourage fungal growth. Regular cleaning using soft brushes, vacuum cleaners with HEPA filters, and lint-free cloths ensures a safe environment. Floors, shelves, and storage furniture should be kept free of clutter to allow proper airflow.
- 2. Pest Management Using Traditional Methods** - While modern integrated

pest management (IPM) is highly effective, traditional Indian practices offer sustainable solutions. For example: Using dried neem leaves or powdered turmeric in storage areas to repel insects.

- Burning camphor periodically to discourage silverfish and bookworms.
 - Placing cloves or dried ginger near stacks to deter pests.
 - These methods, combined with regular monitoring, reduce dependence on harmful chemical fumigants.
 - Saraswati Mahal Library Mixture, developed in the 19th century, is such a traditional method of preventing insects from damaging the collections.
3. Covering and Enclosures - Protecting manuscripts and books from dust, light, and handling damage is vital. Acid-free, lignin-free boxes, folders, and wrappers are ideal for long-term storage. Such enclosures prevent exposure to fluctuating humidity and harmful pollutants. Visual documentation, including images of these protective materials, can help custodians select and use them effectively.
 4. Proper Handling and Movement - Many valuable manuscripts are damaged not by environmental factors but by careless handling. Basic training can instill correct practices such as:
 - Supporting the spine of books when opened.
 - Using clean, dry hands or gloves when handling fragile items.
 - Avoiding stacking heavy volumes on top of delicate manuscripts.
 - Transporting items in cushioned trays or boxes rather than carrying them in bare hands.
 5. Environmental Control - Even without expensive air-conditioning systems, some environmental controls are possible:
 - Avoid storing books and manuscripts against damp walls.
 - Use silica gel or desiccants in closed storage spaces to absorb excess humidity.
 - Keep windows fitted with UV filters or curtains to minimise light damage.
 - Ensure cross-ventilation to reduce fungal growth.
 6. Disaster Preparedness -India's climate makes collections vulnerable to floods, leaks, and fire hazards. Training custodians in emergency response—such as safely drying wet books, or isolating mould-affected materials—can make a critical difference in preventing large-scale losses.

Empowering Custodians Through Training

One of the most compelling arguments for preventive conservation in India is that it empowers custodians to take an active role in preserving their cultural heritage. A well-designed short training programme, perhaps lasting a few days, can equip librarians, archivists, and collection owners with practical skills. Such programmes should be simple, visual, and hands-on, covering:

- Identification of common deterioration signs (e.g., brittle paper, insect holes, mould).
- Cleaning and safe storage techniques.
- Preparation and use of traditional pest repellents.
- Making and using simple acid-free enclosures.
- Record-keeping to track environmental conditions and pest activity.

Integrating Traditional and Modern Approaches

India's manuscript preservation history predates modern conservation science. Ancient repositories developed indigenous systems for safeguarding manuscripts, such as wrapping palm leaves in cloth, oiling them with citronella or neem oil, and storing them in wooden chests. While some traditional methods require modification to meet current conservation ethics, many remain relevant. Combining this inherited wisdom with scientifically proven preventive measures creates an approach that is culturally rooted, cost-effective, and environmentally sustainable.

The digitization of manuscripts and the preparation of their metadata make it easy to access manuscript content. The use of a digitized copy during consultation reduces the unnecessary handling of manuscripts. Digitization, however, does not mean that the original would be neglected. Digital data, however, is a fragile material, and there is no guarantee of its long-term availability. However, physical data, if kept properly away from the vagaries of time and climate by simple methods, ensures the transmission of physical copies from one generation to the next. Preventive Conservation methodologies are the same set of principles that our ancestors used so effectively that some manuscripts have been preserved in mint condition. The choice of raw materials for creating manuscripts, coupled with the proper housing of manuscripts in wooden boxes, ensured that manuscripts reach our generation intact. Now our duty is to generate resources for the proper housing of manuscripts and protect them from neglect and poor storage conditions.

It is essential that the Disaster Risk Management and Reorganization

projects are launched throughout India in repositories, libraries, temples, mosques, gurudwaras, monasteries, and other similar institutions to preserve the collection for future generations. Without a reorganization project, it is challenging to determine where to begin. Re-org will help in devising a methodology for how to start. Once manuscripts are organized, all other actions can follow. These may include preventive conservation, Digitization, and providing proper enclosures, among others. When manuscripts are lying in a disorganized state, it is difficult to do anything. Often, efforts invested are wasted. Leading organizations, such as ICCROM and UNESCO, have developed methodologies and toolkits that can be used for organizing collections and preparing disaster risk management plans.

The Role of Acid-Free and Lignin-Free Storage Materials

Modern archival-quality storage materials—such as acid-free and lignin-free boxes, folders, and sleeves—play a crucial role in the practice of preventive conservation. They help maintain a neutral pH environment, preventing acidic deterioration of paper. In India, where many manuscripts are already on acidic paper or palm leaf, such storage materials can significantly extend their lifespan. Providing custodians with images, designs, and sourcing information for these enclosures encourages wider adoption.

Preventive conservation of books and manuscripts is not merely a professional concern; it is a national responsibility. In a country like India, with an unmatched wealth of written heritage but a severe shortage of trained conservators, preventive conservation stands out as the most practical, scalable, and effective strategy.

By training custodians in simple yet scientifically sound techniques—ranging from good housekeeping and pest control to proper handling and storage—we can drastically reduce the rate of deterioration. Preventive measures do not replace the need for professional remedial conservation, but they buy precious time, often preventing the need for expensive and complex interventions altogether.

The future of India's documentary heritage depends on a collective commitment: institutions, private collectors, and government agencies must prioritise preventive conservation now. If we act with foresight, integrating traditional practices with modern archival science, we can ensure that future generations inherit not faded, crumbling remnants, but vibrant, well-preserved records of our civilization. Over the past four decades, IGNCA has achieved significant milestones in the field of manuscript conservation. IGNCA is conserving the country's wealth of manuscripts at various locations

and would be more than happy to contribute wherever required in the future.

Remedial Conservation of Manuscripts

Deterioration and Conservation of Copper Plate Inscriptions - Copper plate inscriptions, valued for their historical and epigraphic importance, are vulnerable to various forms of corrosion depending on their composition, burial conditions, and exposure after excavation. Understanding the mechanisms of deterioration is essential for effective conservation.

Mechanisms of Deterioration - Copper, due to its low ionization potential, tends to release electrons and form positively charged ions (Cu and Cu^{2+}). On exposure to air, oxidation begins, and a primary layer of cuprite (Cu_2O) forms on the surface. While this layer can act as an initial barrier, defects within cuprite allow continued ionic movement, leading to further deterioration—a process commonly known as dry corrosion. In humid environments, the presence of oxygen and moisture accelerates the corrosion process. Dissolved copper ions migrate to the surface, reacting with environmental agents to form basic carbonates, such as malachite and azurite, or basic chlorides. While carbonates may form uniform, protective patinas, chlorides often initiate aggressive and destructive corrosion.

The burial environment plays a crucial role in determining the formation of corrosion products. In areas rich in CO_2 , green carbonate layers may form, providing protection. In industrial regions, pollutants such as sulphur dioxide (SO_2) and hydrogen sulphide (H_2S) can form brochantite or antlerite, both of which alter the surface appearance and composition of the metal.

Chloride-Induced Damage and Bronze Disease -One of the most serious threats to copper inscriptions is bronze disease, a chloride-induced, cyclic corrosion process. Cuprous chloride (chalcocite), often originating from burial in marine or coastal environments, remains dormant until exposed to moisture and oxygen. It then converts to copper trihydroxychlorides, expanding in volume and causing pitting, cracking, and the formation of powdery, green deposits. If left untreated, this process can completely destroy the metal.

Chloride contamination may also occur from storage in coastal areas, where salt-laden winds contribute to deterioration. Non-traditional forms of bronze disease have been observed in museums located near the sea, where uncontrolled environments accelerate damage.

Sodium-Related Corrosion -Copper plates buried in soils containing sodium salts may develop chalconatronite, a bluish corrosion product. This compound can also result from earlier conservation treatments involving sodium-based chemicals such as sodium hydroxide, sodium sesquicarbonate, or sodium tripolyphosphate. Additionally, exposure to acetic acid vapours from wooden display cases or certain paints can lead to similar blue corrosion layers.

Corrosion in Copper Alloys - Many copper plates are not pure copper but alloys, each with distinct corrosion behaviour.

- Brasses (copper-zinc alloys) are prone to dezincification, especially in chloride-rich, acidic environments.
- High-tin bronzes develop stable, inert tin oxide patinas, which can be protective.
- Leaded bronzes may suffer selective corrosion of the lead phase, resulting in whitish crusts.
- Arsenical and antimonial copper alloys can experience localized corrosion due to elemental segregation during casting.

Conservation Strategies

Conservation of copper plate inscriptions involves both cleaning and stabilisation, with the approach tailored to the specific alloy, condition, and historical value of the object.

Cleaning Methods

Mechanical cleaning—using glass-fibre brushes, dental picks, wooden tools, and fine powders such as chalk—is the most controlled method, as it avoids over-cleaning that could damage inscriptions. Electrolytic reduction can effectively remove corrosion products, but it requires skilled handling to prevent metal loss. Chemical treatments, such as immersion in a sodium sesquicarbonate solution, can remove chlorides; however, the process is slow and may alter the patina. Other agents, including citric acid, EDTA, and alkaline solutions, have been used; however, they may cause surface changes.

Stabilisation Methods

Three primary approaches are used to stabilise copper inscriptions suffering from bronze disease:

1. Complexing agents – Benzotriazole (BTA) is widely used to inhibit

- copper ion activity, though it must be handled with care due to toxicity. Aminomercapto thiazole (AMT) is effective but may discolor the surface.
2. Chloride immobilisation – Traditionally achieved using silver oxide, but more effective results have been obtained by using zinc dust to form protective zinc hydroxide chlorides.
 3. Protective coatings – Microcrystalline wax, Paraloid B-72, and polyvinyl acetate are applied to block moisture and pollutants. Coatings, however, degrade over time and must be periodically removed and reapplied.

Special Considerations for Inscriptions

When conserving copper plate inscriptions, the preservation of legibility is paramount. Stable, non-harmful patinas should be retained to maintain historical authenticity, and cleaning should be limited to the removal of active, damaging corrosion. Long-term preservation requires controlled storage environments with low relative humidity, minimal chloride exposure, and protection from industrial and organic pollutants.

The conservation of copper plate inscriptions demands a careful balance between scientific treatment and respect for historical integrity. Understanding the corrosion processes, selecting appropriate treatments, and maintaining stable environmental conditions are crucial to safeguarding these invaluable historical records for future generations.

Conservation Methodology for Books and Manuscripts

The conservation of books and manuscripts begins with a thorough assessment of their condition to identify the types and extent of deterioration. Paper, the primary material of most manuscripts, is highly sensitive to environmental conditions, physical handling, and chemical changes over time. Fluctuations in temperature and relative humidity can cause expansion and contraction of paper fibres, leading to warping, cockling, or brittleness. High humidity encourages mould growth, while low humidity increases fragility. Exposure to light, particularly ultraviolet radiation, accelerates the degradation of cellulose and the fading of inks or pigments. Pollutants in the environment, such as sulphur dioxide and nitrogen oxides, can react with paper fibres, causing embrittlement and discoloration.

Preventive conservation is the first line of defence. This involves providing a stable environment with controlled temperature (around 18–22°C) and relative

humidity (45–55%), limiting exposure to light, and ensuring adequate air filtration to remove dust and pollutants. Manuscripts should be stored in acid-free boxes, folders, or wrapping materials to buffer against environmental fluctuations and protect from mechanical damage. Shelving should be sturdy, smooth, and free from acidic finishes. In cases where manuscripts are bound, storage should be upright with proper support, while unbound leaves are best kept flat.

When active conservation treatment is necessary, it must be carried out by trained professionals using minimally invasive methods. Dry cleaning is often the first step, using soft brushes or specialised erasers to remove surface dirt. Tears can be repaired with Japanese paper and reversible starch paste. In cases of severe acidity, deacidification treatments, such as aqueous or non-aqueous alkalisng agents, may be employed to neutralise acids and extend the paper's lifespan. Consolidation of flaking inks or pigments requires careful selection of appropriate consolidants, ensuring they do not alter the appearance or chemistry of the manuscript.

For leather or parchment bindings, conservation may involve gentle cleaning, humidification for reshaping, and consolidation of weakened areas with suitable adhesives. Care must be taken to preserve original tooling, decoration, and structure. Biological threats, such as mould or insect infestations, should be addressed through non-invasive methods such as controlled freezing or anoxic treatments.

Ultimately, documentation is a crucial component of the methodology. Every intervention should be recorded, including the materials and techniques used, along with photographic evidence before, during, and after treatment. This ensures transparency, facilitates future research, and supports ongoing preservation. Conservation is an ongoing process, and the long-term safeguarding of books and manuscripts depends on the combination of preventive care, appropriate treatment, and continuous monitoring.

Conclusion

India's written heritage, encompassing manuscripts and copper plate inscriptions, embodies centuries of cultural and intellectual history. Preserving it is not merely a professional responsibility, but a collective national duty.

Given the vastness of collections and the shortage of trained conservators, preventive conservation remains the most practical and scalable approach

to conservation. It empowers custodians, integrates traditional wisdom with modern science, and significantly reduces deterioration risks. Remedial conservation, where necessary, must always respect historical authenticity and adopt minimal, reversible interventions.

Ultimately, the survival of India's documentary heritage depends on proactive action. By combining preventive care, selective remedial treatment, digitization, and training, institutions like IGNCA, along with state and local custodians, can ensure that this heritage endures as a vibrant resource for scholarship and cultural identity for generations to come.



Working Group VI

Theme: Decoding Manuscripts: Pathways to the Indian Knowledge System

Coordinator

Prof. A. B. Shukla
HoD (BVP), IGNCA

Co-coordinator

Ms. Pallavi Prashant Holkar
Director, Ministry of Culture, Govt. of India

Experts

Prof. Om Nath Bimli
Professor & Head, Dept. of Sanskrit, Faculty of Arts, University of Delhi

Dr. Nagaraj Paturi
Vice Chancellor, INDICA

Prof. Amba Kulkarni
Dept of Sanskrit Studies, University of Hyderabad, Hyderabad

Prof. Nachiketa Tiwari
Professor, Dept. of Mechanical Engineering, IIT Kanpur

Prof. Ram Nath Jha
School of Sanskrit and Indic Studies, JNU

Prof. Prasad Joshi
Vice Chancellor (A.), Chairperson, Deccan College, Pune

Dr. Mayank Shekhar
Mission Director, NMCM, IGNCA

Prof. Gauri Mahulikar
Professor & Dean of Faculty, Chinmaya Vishwavidyapeeth (Deemed to be University), Ernakulam, Kerala

Dr. Madhusoodana Pai J.
Assistant Director (Research), In-charge, BVP, IGNCA

Introduction:

Indian Knowledge Systems (IKS) is among the world's most ancient and enduring traditions of learning, encompassing diverse disciplines such as philosophy, mathematics, medicine, astronomy, linguistics, law, and the arts. The manuscript tradition (*pāṇḍulipiparamparā*, which comes under *likhitaparamparā* or written tradition) is deeply rooted in the oral tradition (*śruti/maukhika-paramparā*), and serves as the principal medium through which Bharata's vast intellectual heritage has been preserved. The written record is, in essence, a reflection and extension of the originally oral knowledge systems in practice. Decoding these manuscripts is not merely an exercise in reading ancient texts—it is a profound engagement with the intellectual legacy of a civilization that shaped unique paradigms of inquiry, knowledge, and understanding.

At its core, decoding manuscripts is an attempt to retrieve the original spoken language (*śabda* or *vāk*) that was encoded into written form¹ using various scripts and orthographic conventions on writing materials such as palm leaf, paper, birch bark, etc. This decoding process involves interpreting not only the linguistic content but also the historical, cultural, and phonetic layers embedded in the manuscript tradition. Hence, decoding the manuscripts requires not only an understanding of the scripts and styles used in historical documents, but also an understanding of their language, subject matter, context, and various linguistic principles. Fortunately, in Bharata, we have a rich heritage of such principles discussed in the *śāstras*.

The script serves as a gateway to deeper, more nuanced meanings. Addressing this requires an understanding of what is encoded through the script, how it is encoded, when it was encoded, and for what purpose. The present concept note outlines the evolution of manuscript culture up to the formation of libraries, examines key aspects involved in deciphering manuscripts, and acknowledges how manuscripts serve as vital pathways to the IKS.

Fundamental Dimensions of Manuscript Decoding

It became necessary to encode information first through language as a means of preserving and transmitting knowledge across individuals and generations, and later through scripts as a durable medium to support oral tradition.

In the *Nirukta*, Yāska records a significant observation about the transmission of knowledge:

1 The process of visually representing speech or verbal sounds is often described as “visible speech”—a transcription of verbal energy into visual form.

“*sākṣātkṛtadharmāṇaṣyobabdhūvuh/te’varebhyahāsākṣātkṛtadharmabhyah
upadeśena mantrān samprāduh/ upadeśāya glāyanto’vare bilmagrahaṇāya
imaṃ granthaṃ samāmnāsiṣuh vedaṃ ca vedāṅgāni ca/*”

(*Nirukta*-1.20)

(There existed *Ṛṣis* (seers) who had direct-realization of *dharma* (knowledge) (*sākṣātkṛtadharmāṇah*). From their own inner experience, they transmitted the mantras to others who lacked such realization (*asākṣātkṛtadharmabhyah*), through *upadeśa* (oral instruction). Over time, however, *upadeśa* alone became increasingly difficult to sustain and began to decline (*upadeśāya glāyantaḥ*). To support later generations, who were unable to grasp and retain the mantras purely through oral means (*avare bilma²-grahaṇāya*)—the *Ṛṣis* systematized and transmitted the body of knowledge in the form of a *grantha³* (text), comprising the *Veda* along with the *Vedāṅgas*)

This illustrates two historical necessities: (1) the need to transmit knowledge beyond direct realization, ensuring continuity through *upadeśa* (oral instruction), and (2) the eventual codification of the oral corpus into systematic oral texts.

Language (*Bhāṣā*)

Language plays a crucial role in communication, serving as the primary vehicle for conveying knowledge, culture, and collective memory. Bharata has given such great importance to language that, in the *Vyākaraṇa-darśana-paramparā* (the philosophical side of *vyākaraṇasāstra*), all reality is viewed as a form of language (*śabda*). This language-holistic view of reality is referred to as *śabdabrahmavāda* in *Vyākaraṇadarśana* and is articulated in great detail in Bhartṛhari’s *Vākyapadīya*. Daṇḍin, in his *Kāvyādarśa*, emphatically declares the supreme importance of language, highlighting it as the very foundation upon which expression, knowledge, and literary excellence rest: -

“*idam andhantamaḥ kṛtsnam jāyeta bhuvanatrāyam/*

2 *bilmam bhilmam bhāsanam iti vā/* (*Nirukta*-1.20), here *bilmam* is explained as *bhāsanam* (exposition), meaning the act of clarifying, or making the hidden sense of words intelligible.

3 *granth sandarbhe* (*pāṇinīyadhātupāṭha*- 9.49, 10.373); *sandarbho bandhanam* – (*kṣīratarāṅgiṇī*).

yadi śabdāhvayaṃ jyotirāsaṃsāraṃ na dīpyate//”

(Kāvyaḍarśa-1.3)

(This entire world would become engulfed in complete darkness if the light known as *śabda* (language) did not shine across the cosmos.)

In the Bharatiya tradition, language (*śabda/bhāṣā/vāk*) is understood as a meaningful sequence of sound units (*varṇa-krama*) arising from the speaker’s deliberate intention (*vivakṣā*), articulated as audible sound (*dhvani*)⁴ through the phonetic process involving *sthāna* (place of articulation), *karaṇa* (articulatory organ), and *prayatna* (effort or manner of articulation); and instantly perishing at the moment of utterance, while the physical sound fades, the mental apprehension of the whole word-form (*sphoṭa*) remains in the listener’s cognition enabling understanding, the communicative act resting on the relationship between speaker (*vaktṛ*) and listener (*śrotr*), functioning through the natural faculties of speech and hearing without dependence on any external devices, and attaining effectiveness through clarity, correct articulation, and vivid expression (*vyaktā vāk*)⁵ as emphasised in grammatical, *Mīmāṃsā*, and other texts.

Significance of Oral Tradition in Bharata

Each *rāṣṭra* or civilisation has its own genius/speciality. It is the uniqueness of *Bhārata-rāṣṭra* that its emphasis on oral tradition led to what historians perceive as the absence of written documents, despite having the most ancient oral textual tradition, which is much older than the beginning of the written tradition and is well-preserved.

In traditional texts even today, the expression *śrūyate*, meaning “it is heard,” is used by the *śāstra* scholars. When *kāvyas* are classified, they are divided into *drśya* and *śravya*, not *likhita* or *pāṭhya*. This illustrates the view of *śravyakāvyas* as audible, rather than readable. The use of the words *śravya* and *śrūyate* is found even during the time when traditional written manuscripts became widespread. This establishes that these words do not indicate the

4 The *śabda* (speech) manifests in four levels—*parā*, *paśyantī*, *madhyamā*, and *vaikharī*. *Parā* is the supreme, undifferentiated potential within consciousness, *paśyantī* is the inner perception of meaning, and *madhyamā* is the mental formulation of words, all of which are subtle and natural. *Vaikharī* is the articulated, audible speech, perishable at the moment of utterance and apprehended by the listener. Script (*lipi*) records only this final stage, functioning as an external tool.

5 *bhāṣ vyaktāyāṃ vāci* (*Pāṇinīyadhātupāṭha*–1.696)

absence of written manuscripts, but they only indicate the primacy given to oral transmission of knowledge or poetry over the written transmission.

This oral preservation was not merely a cultural preference, but the foundation of an intellectual heritage, where languages such as Sanskrit, Pali, Prakrit, and Tamil formed the core linguistic bases for diverse schools of philosophy, science, and art within the framework of the IKS.

Script (Lipi)

Bhāṣā (language) is regarded as eternal, subtle, and rooted in consciousness, whereas script (*lipi*) is considered a later development—a functional tool to render language into visible form through conventional signs. Technically, a script is the visual embodiment of language, representing sound units⁶ and enabling speech to be preserved beyond the moment of utterance. As an external apparatus, it requires physical or digital media such as writing materials, surfaces, ink, or modern electronic devices.

A writing system, built upon a script and its governing rules, provides a structured method for visually representing verbal communication. Writing offers a durable and reliable medium for storing and transmitting information across time and space. Script thus emerged to preserve and transmit language, allowing the documentation, codification, and dissemination of knowledge.

Yet, despite this utility, the script never supplanted the oral authority of language. Within this framework, script remains secondary but indispensable, serving as a supportive aid in safeguarding and extending linguistic heritage.⁷

6 The foundation of *lipi-vijñāna* (scriptology) lies in “transcription,” yet this transcription itself is rooted in “speech.” Hence, while scriptology may appear secondary from the perspective of linguistics, it inevitably incorporates a substantial portion of phonology and grapho-phonology, with the latter forming its very cornerstone. Wherever linguistic analysis extends down to the level of phonetic strokes or articulatory impacts, these too become matters of scriptological study. ... In conclusion, although scriptology is distinct from linguistics in some respects, a considerable body of significant knowledge is shared by both disciplines. (Source: *Lipi-Vijñāna aur Nāgarī Lipi (in Hindi)* by Om Prakash Bhatia, Surya Prakashan, Nai Sarak, Delhi)

7 In the contemporary context, the rise of AI-driven text generation, as on today, presents a subtle but significant challenge: it risks shifting attention away from the speaker and the listener, thereby weakening the living dimension of language. To give lesser importance to speaker and listener while the device and tools take much importance is to overlook the very foundation upon which

2.2.1 Limitations of a Script

Decoding manuscripts is deeply influenced by the inherent limitations of script. Unlike the oral tradition, script cannot capture subtle phonetic, melodic, and performative nuances that were central to the recitation and transmission of knowledge in different traditions, making such dimensions irretrievably lost in written records. (Script is only one of the many ways of recording or documenting the spoken expression, after the development of audio and video recording tools, which have expanded the modes of preservation and transmission of knowledge beyond the written word)⁸. The following points highlight the limitations of script—

Loss of Oral Nuance

Scripts cannot fully capture tone, pitch, rhythm, or the melodic accents (*svaras*) of Vedic chanting and everyday language use, nor can they convey the emotional and performative aspects of local storytelling traditions.

Multiple Scripts for the Same Language

Sanskrit, Prakrit, and even local vernaculars were historically written in diverse regional scripts (e.g., Old *Nāgarī*, *Grantha*, *Śāradā*), complicating uniform reading and interpretation.

knowledge, poetry, and *śāstra* have thrived for millennia. Therefore, even as technology advances, it must be recognized that in our tradition, language as spoken and heard retains a higher value than script, for it is through orality that meaning, authority, and continuity are most authentically preserved.

8 As a part of transcription, apart from the use of script, audio recording, which emerged only in the late 19th century with the phonograph, enabled for the first time the preservation of not just words but also intonation, rhythm, stress, and tonal variations of speech. This allowed oral traditions such as Vedic chanting and folk recitation to be documented with far greater accuracy than writing alone could achieve. It thus bridged, to some extent, the gap between oral and written traditions, though very late in human history. With further advancement, video technologies have added another dimension, making it possible to capture both visual and audible aspects together. This integration not only preserves the sound but also the gestures, expressions, and embodied practices that are essential to fully transmitting living traditions, thereby providing a holistic record that neither writing nor audio alone could accomplish.

Manuscript Material Constraints

Space limitations on palm leaf, birch bark, or handmade paper encouraged the use of abbreviations, ligatures, or compressed orthography, which can obscure original linguistic details, especially in dialect-rich content.

Dependency on Reader's Prior Knowledge

Many manuscripts presuppose familiarity with Vedic, grammatical, or local idioms; without oral guidance, texts alone may remain incomplete or misunderstood.

Limitations of Cross-Script Representation

Certain regional sounds resist accurate representation across Bharatiya scripts. For instance, the Malayalam *la* (ॡ) lacks an exact equivalent in other Bharatiya scripts, making traditional transcription and transliteration challenging. Contemporary Pan-Indian Unicode encoding accommodates such characters, ensuring their presence in digital text.

2.2.2 Importance of a Script

Regarding the importance of the script, *Nārada*⁹ beautifully emphasizes that-

nākariṣyad yadi brahmā, likhitaṃ cakṣur uttamam/

tatreyaṃ asya lokasya nābhaviṣyac chubhā gatiḥ// (Nāradaśmṛti-4.70)

(If Brahmā had not created writing, regarded as the highest means for perceiving knowledge, the world would have been deprived of an auspicious path)

lūnapakṣo yathā pakṣī, vṛttihīno yathā dvijaḥ/

śīrobhraṣṭā yathā mālā, parvato dhātuvarjitaḥ//

prabhraṣṭalipi śāstraṃ vā ṛgyajurvisvaram yathā/

svarahīnaṃ yathā sāma, padma-hīnaṃ yathā saraḥ//

(Nāradaśmṛti-2.35.20-21)

⁹ Also see: ‘brahmaṇā vācāṃ varṇasvaravicihñitam’ (*Śukranītiśāra*-2.297); ‘dhātrākṣarāṇi sṛṣṭāni patrārūḍhānyataḥ purā’ (*Bṛhaspati-Smṛti*-1.6.2)

(Just as a bird without feathers, or a *dvija* (*Brāhmin*) without conduct, or a garland which is knotless, or a mountain devoid of minerals; So too is a *śāstra* without script, or the Rig and Yajur Vedas without proper *svaras* (intonation), *Sāmaveda* without *svara* (notes of music), or a lake without lotuses are meaningless) are incomplete, ineffective, and bereft of their true essence.)

The centrality of language and script in human civilization is beautifully captured in the words of Ngawang Samten, who observes:

“Language is the greatest cultural achievement of human beings and forms the foundation of all other cultural creations. However, without the invention of script, language would not have acquired its temporal dimension-its ability to enable communication across the boundaries of time. Without script, language is only a medium of communication among one’s contemporaries, and with script, one can communicate to posterity.”

(Ngawang Samten, Preface, History and Paleography of Kharoshti Script.)

2.2.3 Requirements of a Script

The effectiveness of any script is shaped not only by its symbolic system but also by the practical and cultural conditions of its use. Serving as a bridge between spoken language and its recorded form, a script is inevitably constrained by what writing can or cannot capture. The key aspects include—

Mutual Comprehension

For communication to succeed, both writer and reader¹⁰ must share an understanding of the script and the language it represents.

Medium of Inscription

The choice of material (stone, palm leaf, birch bark, paper, or digital storage) determines not only the process of writing but also its accessibility.

Durability of Record

The medium influences whether a record is preserved for centuries or lost quickly, shaping cultural memory.

2.2.4 Writing Systems in Bharata

The earliest known script is that of the *Sarasvatī-Sindhu* Civilization, evident from thousands of inscribed seals and tablets discovered at Harappan sites, including Mohenjo-daro, Harappa, and Dholavira, which indicate a functional writing system. Later, the *Brāhmī* script emerged before the 3rd century BCE, and was prominently used in the Aśokan edicts, widely regarded as the root of most *Bhāratīya* scripts. References to writing are found in early Buddhist and Jain texts such as the *Lalitavistara* and *Pannavaṇāsutta*, which list various scripts (*lipis*), indicating widespread scriptural awareness and instruction. These references also hint at the use of diverse writing materials—such as palm leaves (*tāla-patra*), birch bark (*bhūrja-patra*), clothing, copper plates, and stones—along with instruments like styluses, iron pens (*lohakīla*), and brushes. This rich tradition of writing and scribes served not only administrative and ritualistic purposes but also became the primary medium for preserving and transmitting Bharata’s vast intellectual and spiritual heritage.

2.2.5 Evidence of Script, Scribes, Scribal Schools, etc.

Evidence of scripts, scribes, and scribal schools provides crucial insights into the cultural and intellectual history of ancient Bharata. Such testimonies not only establish the antiquity and diversity of literacy traditions but also underscore the pivotal role of scribes and schools in preserving and transmitting knowledge across generations.

Definition of a Script

yena tu karmanākṣarāṇi nirvartyante tatkarma lipirityucyate

(*Abhidharmadīpa*-258).

(The action by which syllables (*akṣaras*) come into execution is called ‘*lipi*’, the script.)

Number of Scripts

The following passage from the *Lalitavistara*, records the names of sixty-four scripts, along with references to writing traditions, scribes, ink, and schools of learning:

“*katamāṃ me bho upādhyāya lipiṃ śikṣāpayasi / brāhmīkharoṣṭīpuṣkarasāriṃ aṅgalipim vaṅgalipim ... āsāṃ bho upādhyāya catuṣṣaṣṭīlipīnāṃ katamāṃ tvam śiṣyāpayiṣyasi?*”

(*Lalitavistara*-66)

(O teacher, which script will you teach me? Will it be *Brāhmī*, *Kharoshṭhī*, *Puṣkara*, *Aṅga*, *Vaṅga script* ... etc. O teacher, among these sixty-four scripts, which one will you instruct me in?)

Scribe

Rājaśekhara, in his *Kāvyamīmāṃsā* notes the qualities of a scribe as follows-

‘*Sadaḥsamskāraśuddhyartham sarvabhāṣākuśalaḥ, śīghravāk, cārvaṣaraḥ, iṅgitākāraśvedī, nānālīpījñāḥ, kaviḥ, lākṣaṇikaś ca lekhaḥ syāt*’

(*Kāvyamīmāṃsā*-10)

(For the refinement of the composition, the scribe should be skilled in all languages, quick in speech, adept in elegant letters, skillful in grasping the intended meaning of the expressions, conversant with many scripts, imaginative like a poet, and one who is knowledgeable in the science of poetics)

Scribing as a Profession

Traditional lexicons such as the *Abhidhānaratnamālā* highlight the cultural and professional identity of the scribe, defining him through multiple synonymous terms-

“*Lekhakaḥ syāl lipikaraḥ kāyastho ’kṣarajīvakaḥ*”

(*Abhidhānaratnamālā*-586)

(*Lekhaka* is the one who can scribe the scripts, while the term *kāyastha* denotes whose livelihood is writing.)

Lipiśālā An Institution Dedicated to Writing

References across Buddhist texts such as the *Lalitavistara*, *Divyāvadāna*, *Kalpadrūmāvadānamālā*, *Ratnamālāvadāna*, and *Vimalakīrtinirdeśa* highlight the significance of the *lipiśālā*, a hall or institution dedicated to writing, studying, and preserving manuscripts. The *Lalitavistara* devotes an entire chapter, *Lipiśālāsandarśana-parivarta*, where thousands of students gathered to learn writing and literacy. Some references for *Lipiśālā* are as follows-

*“śāstrāṇi yāni pracalanti manuṣyaloke,
saṁkhyā lipiś ca gaṇanāpi ca dhātutantram/
ye śilpayoga pṛthu laukika aprameyāḥ,
teṣveṣu śikṣitu purā bahukalpakoṭyaḥ//
kiṁ tū janasya anuvartanatām karoti
lipiśālam āgatu suśikṣitu śiṣyaṇārtham/
paripācanārtha bahudāraka agrayāne
anyāmś ca sattvanayutānamṛte vinetum//
.... lipiśālāsandarśanaparivarto daśamaḥ”*

(Lalitavistara-10.1-2)

(All sciences that flourish in the human world—such as reckoning, writing, calculation, and the science of elements—together with the immeasurable applications of arts and crafts, have been diligently studied for innumerable aeons. Rather, it is to guide disciples rightly, to nurture many young learners, and to lead countless beings towards higher understanding The tenth chapter is a visit to the *lipiśālā*)

“tena khalu samayena tasyām lipiśālāyām pañcamātrakadāraka-śatāni lipiṁ śikṣanti”

(Kalpadrumāvadānamālā-310.14)

“At that very time, five hundred young disciples were learning writing in that *lipiśālā*,” showing the collective training of students.

“tadā piṭrā niyukto ’sau lipiśālāmupāgamat”¹¹

(Kalpadrumāvadānamālā-81)

“Then, instructed by his father, he went to the *lipiśālā*,” reflecting

11 *tadā vidyopalabdhārthī lipiśālāmupāviśat/ (Ratnamālāvadāna-13.48):* Then, desiring to obtain knowledge, he entered the *lipiśālā*,” highlighting it as a gateway to higher learning.

parental role in sending children for script learning.

“*sarvalipiśālāsu copasaṃkrāmati dāraḥaparipācanāya...*”

(*Vimalakīrtinirdeśa*-155)

“He went into all *lipiśālās* for the training of disciples,”
emphasizing its role in nurturing and preparing students.

2.2.6 Oral Tradition versus Documentation in IKS

Researchers on the Vedas have noted that several branches of the Vedas and mantras within these branches have become extinct, as they were preserved only in oral form. There is a negating outlook towards the recitation of the Veda based on a written source.¹² Consequently, modern admirers of the IKS deeply lament that, due to the absence of documentation, a vast portion of the IKS has been lost. Nīlakaṇṭha Dīkṣita’s observation in the *Nīlakaṇṭhaviṇayaśaṃpū* highlights the profound disparity between the original abundance of the literary heritage and the relatively limited body of texts that has come down to us today:

“*kati kavayaḥ? kati kṛtayaḥ? kati luptāḥ? kati caranti? kati śithilāḥ?*”

(*Nīlakaṇṭhaviṇayaśaṃpū*)

(How many poets are there? How many works have been created?
How many are lost? How many survive? How many have become
endangered or forgotten?)

The literature we now possess is only an indication or a glimpse of that knowledge, preserved indirectly through related documented texts. However, it is also worth noting that the vital significance of script and documentation was recognized as early as the time of the *Arthaśāstra* by Kautilya, also known as Cāṇakya. He notes:

12 *gītī śīghrī śīrakampī tathā likhitapāṭhakaḥ/ anarthajñō’lpakaṇṭhaśca ṣaḍete pāṭhākādhamāḥ*// (*Pāṇinīyaśikṣā*-7.32): (One who chants in a sing-song manner (*gītī*), one who reads too fast (*śīghrī*), one who shakes his head while reciting (*śīrakampī*), one who merely reads from writing instead of memorization (*likhitapāṭhaka*), one who does not understand the meaning (*anarthajñā*), and one with a feeble or unclear voice (*alpakaṇṭha*) – these six are considered inferior reciters).

“vṛttachaulakarmā lipiṃ saṃkhyānaṃ ca upayujīta”

(*Arthaśāstra*-1.5.7)

(The student who has undergone the *caulakarma* (tonsure) should employ script and numerals)

“saṃsthānām antevāsinaḥ saṃjñālipibhiś cāra-saṃcāraṃ kuryuḥ/”

(*Arthaśāstra*-01.12.11)

(The immediate officers of the institutes of espionage (*saṃsthānām antevāsina*) shall, by making use of signs or writing (*saṃjñālipi*), set their own spies in motion (to ascertain the validity of the information.)

“bālīśyād abhiyoktur vā duḥśrutam durlikhitam pretābhiniveśam vā samīkṣya sākṣipratyayam eva syāt/”

(In cases where the plaintiff proves himself stupid, or where bad hearing (on the part of the witness at the time of the transaction) or bad writing is the cause of difficulty, or where the debtor is dead, the evidence of witnesses alone shall be depended on (*sākṣipratyayameva syāt*).)

These evidences clearly negate the idea that there was reliance on memory alone and emphasizes the critical importance of documentation in administration and institutional continuity.

Relation Between a Language and a Script

Kālidāsa in the *Raghuvamśa* beautifully conveys that when the script is properly grasped, the entirety of literature flows into the mind just as a river merges into the ocean at its estuary. He says:

liper yathāvad grahaṇena vāṇmayam, nadīmukheneva samudram āviśat/

(*Raghuvamśa*-3.28).

The metaphor suggests that script functions as the channel through which the vast expanse of human expression, knowledge, and creativity flows into permanence. Without a script, knowledge remains scattered like independent

streams, but with a script, it attains unity, continuity, and preservation, becoming an ever-flowing ocean of tradition. This illustrates the cultural and intellectual depth of the *Bhāratīya* view of writing as not merely technical but as the very medium through which civilization secures its memory.

2.3.1 Relative Ease in Decoding of Bhāratīya Manuscripts: Phonological Basis of Bhāratīya Scripts

Bhāratīya scripts, particularly scripts derived from *Brāhmī* and *Kharoṣṭhī*, are fundamentally phonological in nature, meaning they are designed to represent the sounds of speech systematically. It organizes consonants and vowels based on articulatory phonetics — from velar to labial sounds — reflecting how and where in the mouth sounds are produced. Phonetic concepts like *hrasva* (short), *dīrgha* (long), *pluta* (prolonged), *alpaprāṇa* (unaspirated), *mahāprāṇa* (aspirated), *ghoṣa* (voiced), *aghoṣa* (unvoiced), *anunāsika* (nasalised), *kāṇṭhya* (velar), *tālavya* (palatal), *mūrdhanya* (cerebral/retroflex), *dantya* (dental), *oṣṭhya* (labial), etc., are incorporated in the organisation of the script symbols in *Bhāratīya* writing systems. It can be clearly visible in Brahmic scripts (scripts derived from Brahmi) and *Kharoṣṭhī* script¹³. The features of these scripts were developed based solely on the phonological knowledge of the time. This sound unit-based design ensured that while reading or reciting from the script, the same intended sounds were consistently reproduced.

Invention of Mātrās (Vowel Markers)

Mātrās were developed to indicate vowel length and quality, systematically distinguishing *hrasva* and *dīrgha* vowels in writing, reflecting precise phonetic understanding.

Separate Grapheme for Vyañjanas (Consonants)

Consonants are represented by distinct symbols categorised according to their place of articulation in the mouth, such as *kāṇṭhya*, *tālavya*, *mūrdhanya*, *dantya*, *oṣṭhya*, *anunāsika*, etc.¹⁴

13 The phonological nature of the *Kharoṣṭhī* script challenges several theories propagated by some palaeographers regarding its Aramaic origin.

14 W.S. Allen notes- “Only in the latter part of the nineteenth century, under the influence of Indian teaching, does the recognition of the voicing process make headway.”- (p.37, W.S. Allen, *Phonetics in Ancient India*, Oxford University Press, 1953)

A System of Combining Glyphs to Represent Samyuktākṣaras (Clusters)

Scripts incorporate methods to combine individual consonant symbols to form *samyuktākṣaras* (clusters), such as स्न, त्स्न, त्स्न्य, and स्त्स्न्य, enabling the accurate representation of speech clusters.

Concept of Script Symbols for Specific Samyuktākṣaras (Clusters)

Certain frequently occurring consonant clusters, such as क्ष, त्र, ज्ञ, and श्री, have dedicated symbols to simplify writing and reading.

Special Symbols and Notation

Specific script adaptations and notations were devised to capture the precise sounds, intonations, and aspects of Vedic chants and music, preserving oral traditions in written form.

One-to-One Mapping

The Brahmic scripts exhibit a one-to-one mapping between sounds and symbols, meaning each sound unit in the language has a corresponding grapheme in the script. This property enables the precise representation of spoken language, making Brahmic scripts highly phonological and efficient for recording the many languages of the Indian subcontinent.

In all writing systems that trace their origin to the Phoenician script, each script symbol or letter is assigned a distinct name that differs from the actual sound it represents. For example, the letter has one conventional name, while its phonetic value may be entirely different, and in many cases, a single script symbol may represent multiple sounds depending on its position or surrounding letters. This makes the process of learning such scripts more complex, as the learner is burdened with the additional task of memorising spellings in addition to understanding the sounds.

Phonological Precision in Brāhmī

In Brahmic scripts, a script symbol does not have a separate or arbitrary name at all; rather, the name of the script symbol is identical to the sound it denotes. Moreover, each Brahmic symbol consistently represents only one sound, irrespective of its position or context, making the script inherently phonological, precise, and transparent in its organisation. As a result, in *Brāhmī* and its derivative scripts, the learner is free from the extra burden of memorising spellings, since the writing directly corresponds to the spoken sounds.

Interchangeability of Scripts –

The Brahmic scripts allow their symbols to be adapted or interchanged across different writing systems while preserving phonological values. This flexibility enabled the evolution of numerous scripts derived from *Brāhmī* across the Indian subcontinent; for instance, *sūtre maṇigaṇā iva* (beads on a string), its fundamental principles remained unchanged, only adapted as needed.

| संज्ञा | स्वराः | | | व्यञ्जनानि | | | | | | | | अद्योगवाहः | शिक्षासूत्राणि | |
|-------------|---------|--------|------|------------|---|---|--------|---|---|-----|----|------------|----------------|---|
| | ह्रस्वः | दीर्घः | लृट् | | | | | | | | | | | |
| स्थानम् | | | | | | | नासिका | | | | | | | |
| कण्ठः | अ | आ | आ३ | क | ख | ग | घ | ङ | | ह | ः | <== | 1 | अकुहविसर्जनीयाः कण्ठ्याः (Velar) |
| तालु | इ | ई | ई३ | च | छ | ज | झ | ञ | य | श | | <== | 2 | इचुयशस्तालव्याः (Palatal) |
| मूर्धा | ऋ | ॠ | ऋ३ | ट | ठ | ड | ढ | ण | र | ष | | <== | 3 | ऋटुरषा मूर्धन्याः (Retroflex) |
| दन्ताः | लृ | - | लृ३ | त | थ | द | ध | न | ल | स | | <== | 4 | लृतुलसा दन्त्याः। (Dental) |
| ओष्ठौ | उ | ऊ | ऊ३ | प | फ | ब | भ | म | | | ५प | <== | 5 | उपूषध्मानीया ओष्ठ्याः। (Bilabial) |
| कण्ठतालुः | - | ए | ए३ | | | | | | | | | <== | 6 | ए ऐ कण्ठतालव्यौ। (Palato-velar) |
| कण्ठोष्ठम् | - | ओ | ओ३ | | | | | | | | | <== | 7 | ओ औ कण्ठोष्ठ्यौ। (Palato-labial) |
| नासिका | | | | | | | | | | | ं | <== | 8 | अनुस्वारयमा नासिक्याः (Nasal) |
| दन्तोष्ठः | | | | | | | | | व | | | <== | 9 | वकारो दन्तोष्ठ्यः। (Labio-dental) |
| जिह्वामूलम् | | | | | | | | | | | ५क | | 10 | जिह्वामूलीयो जिह्व्यः
(sounds from Root of the tongue) |
| कुल | 5 | 8 | 9 | 25 | | | | 4 | 4 | 4+4 | 63 | | | (तत्र स्थानकरणप्रत्ययेभ्यो वर्णास्त्रिषष्टिः) |

Manuscripts Heritage

Since manuscripts, as tangible embodiments of language through script, any problem in either domain reflects directly in the document's readability and comprehension. For instance, a word written in an archaic form or in a script variant not in common use today may pose obstacles for the modern reader or scholar. Additionally, a single script may represent multiple languages, and a single language may appear in different scripts across time and regions, further complicating the decoding process.

2.4.1 From the Invention of Script to the Journey of Manuscript Libraries

The evolution of writing from its earliest conception to the establishment of libraries reflects both linguistic depth and cultural expansion. The very Sanskrit root *lip/lib*¹⁵ (*akṣaravinyāse*) carries the meaning of arranging or inscribing letters, which later extended into multiple senses such as smearing, anointing, covering, or engraving. This semantic richness demonstrates how the idea of marking or imprinting—whether on surfaces, walls, or minds—developed into the notion of writing. The English word “library,” derived from the Latin *libraria* (from *liber*), interestingly relates to the root *lip/lib* in Sanskrit¹⁶. Thus, script and libraries share a deep etymological and cultural bond: script refers to the act of inscribing, and library refers to the repository where inscriptions are stored and preserved.

2.4.2 Purpose of Manuscripts and a Library

Ranganathan’s five laws of library science, though articulated in the context of modern libraries, carry deep relevance for manuscript heritage. When applied to manuscripts¹⁷ These principles highlight their role not as static relics but as dynamic sources of knowledge, awaiting interaction with readers and scholars. They also emphasise the need for accessibility, preservation, and growth to ensure manuscripts remain integral to contemporary intellectual life.

Manuscripts are for Use

Manuscripts should not be locked away or treated merely as museum

15 As seen from the root *lip-upadehe*, in usages such as *lipta* and *limpati*, and in scriptural references like *nā māṃ karmāṇi limpanti* (Bhagavadgītā-4.14) or poetic attestations such as *lipteṣu bhāsā grhadehalīnām* (*Śiśupālavadham*-3.48), this semantic richness shows how the simple notion of marking or imprinting—on surfaces, walls, or minds—gradually evolved into the abstract concept. This sense survives in Marathi *limpane*, meaning to plaster or cover a wall. Early awareness of script as a codified system is further evident in ‘*indravaruṇa-bhava-śarvarudra-mṛḍa-himāranya-yavaya-vanamātulācāryāṇāmānuk*’ (*Aṣṭādhyāyī*-4.1.49) and its explanation in the *Kātyāyana-vārttikam*- ‘*yavanāt lipyām*’ and in *kāśikā*- ‘*yavanānī lipiḥ*’.

16 Bhat, Vasantakumar. *Pāṇḍulipi evaṃ Samīkṣit Pātha-Sampādana*, Indira Gandhi National Centre for the Arts, New Delhi, 2025, p. 6.

17 Manuscripts naturally fall within Ranganathan’s concept of “book”, understood as any carrier of recorded knowledge.

objects; they are meant to be studied, interpreted, and transmitted. They are living carriers of an intellectual legacy that must remain accessible to scholars and society.

Every Reader has a Manuscript

Each scholar or reader has a manuscript relevant to their pursuit. The wide range of subjects—philosophy, science, literature, ritual, the arts, and more—ensures that manuscripts can meet the diverse intellectual needs of researchers.

Every Manuscript has its Reader

No manuscript is insignificant; even obscure or damaged works await a suitable reader, editor, or interpreter who can rediscover their value and bring their knowledge to light.

Save the Time of the Reader

To facilitate effective scholarly engagement, the time of the manuscript reader, editor, or searcher must be saved. This is possible through accurate cataloguing, metadata tagging, digital access, and subject-based retrieval tools.

The Manuscript Repository is a Growing Organism

Manuscript collections are dynamic entities. They must continuously evolve through acquisition, preservation, digitisation, and collaboration with scholars, so that ancient texts remain active participants in contemporary intellectual discourse.

2.4.3 Manuscripts Between Reverence and Neglect

A popular verse warns us in the voice of the manuscript itself, lamenting the dangers it faces if not preserved with care.

tailād rakṣa jalād rakṣa rakṣa māṃ ślatha-bandhanāt/

mūrkhahaste na dātavyam¹⁸ iti roditi pustakam//

The mention of *taila* (oil) is *upalakṣaṇa*, signifying not just protection from stains but from fire and other destructive elements. *Jala* warns against water, dampness, and decay that silently erode manuscripts. *Ślatha-bandhana* highlights the dangers of loose bindings and neglected upkeep, warning us that without proper conservation, cataloguing, and systematic preservation, texts

18 *parahastagatād rakṣa.*

are lost. Most severely, the manuscript laments being placed in *mūrkhā-hasta*, the hands of the ignorant.

Beliefs around holding a manuscript with an individual or a house have both positive and negative consequences. Beliefs in the beneficial sacred power of the manuscripts helped preserve them, while fears of harmful supernatural powers from the manuscripts led to their destruction. These contrasting beliefs shaped the fate of manuscripts—where reverence ensured their protection and transmission, while fear led to their burning, washing away in rivers, or deliberate neglect. Thus, manuscripts became both sanctified treasures and vulnerable objects, their survival hinging not only on material conditions but also on human outlook.

The warning is clear: manuscripts must be preserved with vigilance, both materially and intellectually, lest they perish or be misused, taking with them irreplaceable knowledge.

2.4.4 Present Situation of Manuscripts

The ultimate purpose of a manuscript is to preserve and transmit knowledge across generations. It serves as a vital bridge connecting the wisdom of the past with the intellectual needs of the future. However, the current situation is deeply concerning—scholars with expertise in various scripts and subjects often struggle to access manuscripts with ease due to the above-mentioned *mūrkhā-hasta-paramparā*. This disconnect hampers the continuity of our knowledge traditions and risks severing the living link between ancient insights and contemporary scholarship.

Despite their importance, many manuscripts remain uncatalogued, untranslated, or physically endangered. The absence of trained manuscriptologists, inadequate funding, and fragmentation across libraries pose serious hurdles. However, initiatives by the National Mission for Manuscripts, Indira Gandhi National Centre for the Arts (IGNCA), and several emerging digital archives across the country are beginning to address these gaps.

Moreover, the integration of manuscript studies into mainstream academia, public exhibitions, and community engagement has the potential to democratize access to these treasures. Cross-disciplinary collaborations among scientists, linguists, and technologists can breathe new life into ancient wisdom.

Decoding manuscripts

The decoding of manuscripts can broadly be viewed in two stages. The first spans reproduction to transliteration, encompassing collection,

cataloging, conservation, digitization, transcription, and translation, which together preserve and render the text accessible. The second stage involves interpretation, where the recovered text is critically examined, contextualized, and analyzed in terms of its intellectual, cultural, and historical significance.

Manuscript Reproduction up to Transliteration

Reproduction of manuscripts in Bharata has a long and continuous tradition. In ancient times, manuscripts older than approximately three hundred years were generally recopied within libraries to safeguard them from the perishable nature of writing materials such as palm leaves and birch bark, thereby ensuring the continuity of knowledge across generations. This vital work was carried out by the *lipikaras*, who were not mere scribes but trained scholars dedicated to the profession of manuscript reproduction. The manuscript copies, often called *putrikās*, maintained a high degree of accuracy owing to the commitment, skill, and discipline of the *lipikaras*. Such remarkable precision greatly reduced the otherwise arduous task of collation, for while an increase in the number of manuscripts with variant readings complicates the work of textual criticism, the accuracy of these traditional copies ensured relative uniformity and reliability.

With the advent of printing technology and later digital tools, the methods of reproduction evolved significantly, yet the underlying concern remained the same—preserving knowledge for posterity. Digital texts have introduced the advantage of searchability and rapid access, but this has often come at the cost of accuracy, creating new challenges in safeguarding the textual integrity of manuscripts. Modern technologies, such as Optical Character Recognition (OCR), have further enhanced the ease of access and reproduction; however, the accuracy of text production without human intervention, unlike the meticulous work of the *lipikaras*, remains a serious concern.

Currently, the reproduction of manuscripts, from their collection to their transliteration, forms the foundational stage of engaging with the knowledge heritage. This process begins with sourcing manuscripts, followed by systematic cataloguing and conservation. Once preserved in this manner, manuscripts are transcribed and transliterated across scripts to broaden readership. Together, these stages safeguard fragile originals, provide reliable reproductions, and prepare the ground for accurate interpretation and translation. Thus, these steps constitute not only a base for interpretation but also the very point at which interpretation begins.

3.1.1 Collection

Manuscripts are sourced from diverse repositories, including libraries, archives, private family holdings, temples, monasteries, and through dedicated field surveys. This stage requires careful negotiation, documentation, and respect for community custodianship, ensuring that valuable materials are systematically identified and brought into the purview of preservation and study.

3.1.2 Cataloguing

Once collected, manuscripts are cataloged with essential metadata, including title, author, script, language, subject matter, folio count, and physical condition. Cataloguing creates an organised record, enabling accessibility for researchers by reducing the time and providing the first layer of scholarly apparatus for further study and critical editions.

3.1.3 Conservation

Conservation involves both preventive and curative measures. Preventive conservation ensures proper storage, handling, and environmental conditions to slow deterioration, while curative conservation addresses existing damage from ink corrosion, insect infestation, brittleness, or wear. This step safeguards fragile manuscripts for future generations.

3.1.4 Digitization

Digitisation converts manuscripts into high-resolution digital images, creating a permanent, shareable record. It minimises physical handling of the fragile originals while allowing scholars worldwide to access and study the material. Digitisation also opens the door to computational tools such as OCR, script analysis, and digital catalogues.

3.1.5 Transcription

The historical shift in mediums—from stone inscriptions and copper plates to palm leaves, birch, handmade paper, printed books, and now digital repositories—highlights the constant need to adapt technologies of preservation¹⁹. In this continuum, transcription plays a vital role as it

19 That which doesn't fit the standards is in danger of being left behind. History has shown that in previous media transitions the knowledge that fails to adapt to the new medium recedes from public view to the restricted domain of the endeavoring antiquarian research scholar or be-

reproduces the manuscript's content into a human-readable and editable format in the very script in which it was originally composed. This process requires palaeographic skills to decipher faded or damaged letters, recognise scribal variations, and accurately render ligatures, orthographic conventions, and diacritical marks. Thus, establishing transcription in the present standard is an indispensable task, ensuring that fragile and time-sensitive manuscripts are safeguarded, accessible, and meaningfully preserved for future generations.

3.1.6 Transliteration

Rendering the text from its original script into another script without altering the language enables readers unfamiliar with the original script to access the content. This step demands consistent transliteration schemes (e.g., IAST for Indic scripts) and careful handling of ambiguous or context-dependent characters. This task requires mastery not only of the script but also of the underlying language, grammar, and literary style, as well as subject expertise in the domain to which the text belongs. Only through this combination of palaeographic skill and intellectual context can the critic produce a transliteration that is faithful, consistent, and reliable for further translation and interpretation.

Interpretation Stages

Each manuscript carries meaning at multiple levels—script, language, genre, and cultural context—which must be carefully unpacked to arrive at a fuller understanding. The interpretive stages begin with recognising the script and textual form, move through linguistic and philological analysis, and extend into genre-specific readings. Ultimately, these insights are situated within their historical, intellectual, and cultural contexts, enabling the manuscript to reveal its deeper significance. This staged approach ensures that a manuscript is not only transcribed but truly understood in its original intent and function.

3.2.1 Aspects to Consider in Interpreting Manuscripts

Decoding a manuscript is a scholarly discipline involving reading the script, identifying the language, breaking the expressions into linguistic units such as sentences, clauses, phrases, words etc. or lines of a metrical verse etc., identifying the intended meaning of the expressions with a discerning knowledge of the meanings as per the historical time of composition, subject matter or the field covered and its technical terms, poetical sensibilities, author's style, while taking care of resolving the problems arising from the comes irretrievably lost. (Peter Scharf, *Linguistic Issues in Encoding Sanskrit*)

damage to the manuscripts, scribal errors, manuscripts scribing traditions, medium based limitations etc.

From Where / Its Source?

Tracing the origin of the encoded material, which could involve identifying the author, scribe, school of thought, or tradition from which the content emerged, is a crucial task.

Misattribution in Catalogues: In many Descriptive Catalogues and manuscript lists, however, instead of the actual *granthakartā* (author – one who composes the text), the names of the *vyākhyātā* (commentator – one who explains the original text), *lipikara* (scribe – one who copies the manuscript), *svāmī* (owner – one who possesses the manuscript), *pāṭhaka* (reader – one who studies it), or *saṃśodhaka* (editor – one who revises it) are often carelessly entered.

Role of Scribes and Scribal Errors: Because writing or scribing was a specialised skill typically held by artisans such as stone sculptors, metal sculptors, or those skilled in using a metal stylus on palm leaves or other writing tools on various materials, this skill necessarily rested with the experts. In the register of language used for the texts or their content, experts dictated the content to scribes, leading to a gap between the intended and the written. This is what is viewed as scribal errors in manuscript research.

What is Encoded?

Identifying the nature of the information concealed within the manuscript, which could include language, symbolic systems, scientific data, philosophical concepts, ritual instructions, or layered allegorical meanings.

Technical Vocabulary and Shifting Meanings Across Traditions: Decoding manuscripts requires careful attention to the technical vocabulary and conceptual frameworks of the school of thought or tradition in which the text was produced. Every intellectual lineage develops its own precise set of terms (*paribhāṣā*), often with meanings that may not be obvious outside that context. A word that seems familiar in one discipline can carry a very different meaning in another. For example, the term *prakṛti* in *Sāṅkhya* philosophy refers to primordial matter, the root cause of the manifest universe, while in *Vyākaraṇa* (grammar) it denotes the base form of a word to which affixes are added. Without awareness of such context-specific usages, a reader or editor may mistakenly interpret the text, leading to a distortion of its intended meaning. Thus, situating a manuscript within its proper philosophical or disciplinary background is essential not only for accurate translation but also for grasping the depth of encoded knowledge.

How Much Information is Encoded?

It is necessary to first determine how much information is lost during the encoding process before assessing the amount of information actually encoded using a script. This requires evaluating the extent of loss that may have occurred in transcribing an oral tradition into written form. The absence, or rare occurrence, of punctuation marks used in the present age further complicates the situation in manuscript interpretation. Furthermore, the act of transcription can introduce omissions, alterations, or standardisations—whether deliberate or inadvertent—that diminish the richness of the original oral content. Therefore, a thorough decoding must account for both the dimensions of information that have been preserved and those that may have been lost. Once this is established, it becomes possible to assess the volume and density of information actually encoded, as some manuscripts may contain multiple layers of meaning, marginalia, hidden acrostics, or embedded diagrams beyond the main text.

When it was encoded

The process of understanding the content as it was originally intended requires reconstructing the cultural, linguistic, and intellectual context in which the manuscript was created, so that modern interpretation aligns as closely as possible with the meaning intended at the time of encoding. In this, *deśa* (geographical–cultural setting) and *kāla* (historical period) play a crucial role, as variations in regional traditions, dialects, socio-political conditions, and prevailing scholarly paradigms can significantly influence both the form of expression and the interpretive framework of the text. Without situating the manuscript in its correct *deśa–kāla* context, decoding risks imposing anachronistic or culturally disconnected interpretations.

Where the Encoding Resides

Locating the specific areas in the manuscript where encoded content is embedded—whether in the main script, marginal notes, illustrations, pagination, or even in the choice of materials, inks, format, etc.

How has the encoding been Carried Out?

Analysing the techniques, stylistic conventions, or intentional distortions used to conceal, protect, or structure the content, such as deliberate misspellings, substitution of characters, or layered commentaries.

Only by systematically addressing each of these aspects can the manuscript be decoded with accuracy, revealing not just its surface script but the full

spectrum of meaning embedded within it. In this light, Textual Criticism becomes an essential stage, providing the scholarly tools to evaluate variants, establish reliable texts, and trace the history of transmission with methodological precision.

3.2.2 Textual Criticism

The scholarly comparison of different manuscript witnesses to identify scribal errors, omissions, interpolations, and variant readings. The goal is to reconstruct the most authentic form of the text, often through stemmatic analysis, collation, and the use of critical apparatus. Identifying the *mūlapāṭha* (original text) and distinguishing it from later *apapāṭha* (interpolations) often requires careful attention to textual markers such as the *puṣpikā*²⁰ and *uttarapūṣpikā*²¹.

3.2.3 Translation

Rendering the content into another language while preserving both meaning and stylistic nuances requires subject expertise, cultural sensitivity, and awareness of idiomatic expressions, as literal translation may distort the original intent.

Thus, translation helps to retain the *mūlapāṭha* (original text) meaningfully by preserving not only words but also the intent, context, and layered significance of the original. In this way, it serves both as preservation and interpretation, keeping the integrity of the source intact while making it accessible across languages and cultures.

20 In relation to a 'Text' (such as the title of the work, chapter sequence, beginning portion like invocations or praises, concluding portion indicating auspiciousness or peace, date and place of writing, patron, subject matter and order of study, relationship and purpose, reason for composition such as praise of the benefactor, and the textual status), or in relation to the author (such as the author's name, lineage, and period of life), any reference recorded by the writer (whether the author himself, a disciple, sub-disciple, commentator, or scribe) in any part of the text (beginning, middle, or end) is can be called a *puṣpikā* (colophon).

21 The passage distinct from the 'Main Text', recorded by the *lipikāra* (scribe) and providing details such as the period of writing, place, and reason for writing, is called *uttara-puṣpikā*, and it serves to establish the identity of the writer. These notes play a vital role in preserving the integrity of the text.

3.2.4 Annotation

Adding explanatory notes, glosses, and references that clarify obscure words, technical terms, cultural allusions, or intertextual references. Annotations bridge the gap between the original context and the modern reader's understanding.

3.2.5 Contextualization

Situating the text within its historical, geographical, cultural, and intellectual framework. This involves recognizing its place in broader traditions, schools of thought, or genres, and understanding the *deśa-kāla-pātra* (place-time-audience) factors that shape its composition.

3.2.6 Comparative Study

Analysing the manuscript alongside related texts within the same tradition or across traditions. Comparative work may highlight influences, shared motifs, divergent interpretations, or parallel developments in different cultural milieus.

3.2.7 Publication

Preparing a scholarly edition that integrates the critical text, apparatus, annotations, translations, and interpretive essays. Publication makes the decoded manuscript available to both academic audiences and, where appropriate, the wider public.

3.2.8 Dissemination

Sharing the outcomes of decoding through lectures, workshops, research articles/dissertations, online repositories, exhibitions, and educational resources. Dissemination ensures that the recovered knowledge re-enters the living stream of the IKS and contributes to contemporary intellectual life.

Genre-wise Decoding of Manuscripts

Decoding manuscripts is not a uniform process but one that varies according to the genre of the text being studied. Each genre presents specific challenges in transliteration and interpretation, requiring different skills, peer review mechanisms, and integration of sources. The following subsections outline these genre-specific considerations.

3.3.1 Creative / Literary Texts

Manuscripts of *gadya*, *padya*, *campū*, *nāṭaka*, and other *kāvya* categories are rich in stylistic ornamentation but pose unique decoding challenges. The ambiguity of poetic meters, deliberate wordplay, and intertextual allusions often require the editor to be both philologist and literary critic. Even minor scribal errors—such as confusion between similar letters—can disturb the rhythm of a verse or obscure a pun. Critical editions of such texts must therefore go beyond mechanical reproduction, demanding a sensitivity to aesthetics, prosody, and the literary conventions of the period.

3.3.2 Śāstra Texts

Decoding *śāstra* manuscripts—whether on grammar, logic, poetics, astronomy, medicine, or *dharmaśāstra*—requires subject expertise in addition to linguistic competence. These works are typically technical, dense, and replete with terminological precision. A single misread syllable can alter the interpretation of a grammatical rule, a medical formulation, or an astronomical calculation. Transliteration in this genre requires deep familiarity with technical vocabulary, while interpretation requires comparison with parallel traditions, commentaries, and existing critical editions. Hence, collaboration between linguists, domain specialists, and historians of science is indispensable in editing *śāstra* manuscripts.

3.3.3 Mystic / Esoteric Texts

Manuscripts on mantra, tantra, yoga, alchemy, and occult practices often circulate in fragmentary or secretive traditions. Their transmission history is marked by intentional obscurity, characterized by cryptic language, coded diagrams, or symbolic abbreviations. Decoding such texts demands more than philology; it requires knowledge of ritual practice, esoteric symbolism, and sometimes even oral traditions of recitation or initiation. Moreover, these texts often vary drastically between manuscripts, reflecting localized or sectarian innovations. Peer review here must include practitioners or scholars conversant with esoteric traditions.

3.3.4 Religious Practice Texts

Ritual manuals (*paddhatis*), *stotras*, *pūjā-vidhi* texts, and *vrata-kathās* are often preserved in manuscripts intended for functional rather than literary use. This means they are prone to scribal simplifications, omissions, or adaptations to local customs and practices. Decoding such texts requires

recognizing the interplay between the normative scriptural tradition and its lived ritual application. Transliteration and translation must therefore be handled with care, distinguishing between local variations and Pan-Indian conventions. Integrating multiple manuscripts helps recover not only the “text” as an abstract entity but also the diverse ways in which it was performed and transmitted in practice.

3.3.5 Biographical and Historical Texts

Genres such as *vaṃśāvalī* (genealogies), *sthala-purāṇa* (local legends), *rājacaritra* (royal chronicles), and *jīvanacarita* (biographies of saints, scholars, or rulers) are valuable historical sources, but they also raise distinctive critical issues. These manuscripts often mix fact with myth, chronological records with hagiographical embellishments. Decoding them involves historical analysis to verify dates and events, as well as comparative work across multiple manuscripts to detect interpolations. In these cases, the critical edition does not aim only to recover the author’s original text but also to map layers of historical memory, editorial accretions, and regional perspectives preserved in different versions.

3.3.6 Peer Review Mechanisms

Decoding is prone to human error at every stage, from misreading damaged letters to mistranslating obscure terms. A structured peer review process is therefore indispensable. Peer review of transcriptions helps verify paleographic accuracy, while review of transliterations ensures that the transfer across scripts remains consistent and intelligible. Similarly, translations and annotations must be reviewed by subject experts, as mastery of language alone may not be sufficient without contextual knowledge. Peer review, in this sense, serves as a safeguard against personal bias or oversight, allowing the scholarly community, rather than a single individual, to determine what constitutes a “published” or reliable version of a text.

Problems in Decoding Manuscripts

3.4.1 Material and Physical Challenges

Physical Deterioration

Manuscripts on palm leaf, birch bark, or handmade paper are highly perishable. Natural ageing, ink fading, brittleness, insect damage, and fungal growth gradually erase content, often beyond recovery.

Illegible Handwriting

Individual scribal styles, dense ligatures, faded ink, and multiple recopying layers often make manuscripts hard to read even for experts.

Obsolete Scripts

Scripts such as *Śāradā*, *Grantha*, and *Modi* are no longer in general use, and few specialists remain to read them, leaving large portions of manuscripts inaccessible.

Missing Portions

Tears, decay, or incomplete transcription have resulted in missing folios. These losses disrupt continuity, compelling scholars to engage in speculative reconstruction.

Scribal Errors

Copying mistakes, such as omissions, duplications, and transpositions, became embedded in traditions, complicating the establishment of critical editions.

Digitization Causing Physical Harm

Improper digitization techniques—such as untrained handling, harsh lighting, or lamination—can sometimes cause irreversible damage to fragile folios.

3.4.2 Linguistic and Interpretive Challenges

Language Barriers

Texts employ classical languages, medieval vernaculars, or mixed registers, necessitating a profound understanding of both cultural and linguistic contexts. Without this, accurate interpretation is difficult.

Misinterpretation

Inadequate contextual or historical understanding often leads to mistranslations, which distort meaning and misrepresent traditions.

Lack of Standardization

Historical manuscripts exhibit fluid orthography, inconsistent sandhi, and the absence of punctuation, resulting in multiple possible readings.

Untranslated Manuscripts and Prioritization Issues

Large portions of manuscripts remain untranslated, with no systematic framework for prioritization. This leaves significant cultural texts unexplored.

Cultural Sensitivities in Preservation

For many communities, manuscripts are sacred objects. Digitization or removal from temples can be seen as intrusive unless approached with cultural sensitivity.

3.4.3 Infrastructural and Institutional Challenges

Inadequate Cataloguing

Many collections are poorly catalogued, outdated, or inaccessible, making it difficult for scholars to locate and study manuscripts.

Limited Access

Institutional restrictions, fragile conditions, and remote or dispersed repositories create serious barriers to scholarly access.

Weak Institutional Practices and Infrastructure Gaps

Lack of conservation policies, inadequate climate control, and the use of harmful chemical treatments expose manuscripts to further deterioration.

Underfunding and Limited Interdisciplinary Collaboration

Preservation requires collaboration among linguists, historians, and technologists; however, inadequate funding and poor cooperation often hinder its effectiveness.

3.4.4 Technological and Resource Challenges

Poor Digitization Quality

Low-resolution scans, colour distortion, and incomplete imaging reduce the scholarly value of digitized manuscripts.

Resource Constraints

Preservation and digitization require sustained funding, trained staff, and adequate infrastructure, but institutions often lack these essential resources.

Lack of Expertise

The number of trained manuscriptologists, palaeographers, and codicologists is declining, risking the loss of critical interpretive skills.

Fragmentation of Sources

Manuscripts of a single work are scattered across archives, libraries, and private collections, often spanning multiple countries, making integration and

comparative study challenging.

Solutions for Manuscript Preservation and Decoding

3.5.1 Physical Preservation and Institutional Support

Conservation Techniques

Preventive conservation measures such as climate-controlled storage, pest management, and careful handling protocols are essential. Advanced methods, such as deacidification, oiling palm-leaf folios, and the use of specialized housing materials, can slow down deterioration.

Funding and Institutional Support

Sustained funding through grants, fellowships, and infrastructure investment strengthens long-term preservation and study.

Expert Collaboration

Decoding benefits from interdisciplinary collaboration—historians, linguists, scientists, technologists, and community custodians working together.

3.5.2 Decoding and Scholarly Methods

Palaeographic Analysis

A systematic study of historical scripts enables the accurate identification of letter forms and orthographic conventions, which in turn helps to date manuscripts and determine their provenance. Script databases can accelerate decoding efforts.

Script Training Programs

Regular training in ancient and medieval scripts (*Śāradā*, *Grantha*, *Modi*, *Nāgarī*, etc.) ensures the intergenerational transfer of palaeographic expertise.

Linguistic Research

Comprehensive studies of classical and regional languages facilitate translation and interpretation, enabling the reliable reconstruction of meaning.

Critical Edition Methods

Collating multiple copies of the same work helps reconstruct authoritative versions, identify scribal errors, and restore missing portions.

Cross-Referencing Manuscripts

Using parallel manuscripts, commentaries, and related texts provides a broader base for interpreting difficult or damaged passages.

3.5.3 Cataloguing and Access

Standard Cataloguing Protocols

Centralized and recognized cataloguing standards should be adopted, with digital catalogues that are searchable, interoperable, and regularly updated.

Improved Archival Access

Greater public and scholarly access through reading rooms, exhibitions, and digital platforms democratizes manuscript research.

Searchable and Accessible Texts

Structured encoding (e.g., TEI) ensures manuscripts are not image-only but fully searchable and analyzable.

3.5.4 Technological Innovations

Advanced Imaging Technologies

High-resolution imaging, multispectral photography, and scanning reveal faded or hidden text, allowing for non-invasive study.

Use of AI and OCR Tools

AI and OCR adapted for Indic scripts can automate transcription and improve recognition accuracy when trained on historical samples.

OCR Development

Beyond Devanagari, OCR tools should be developed for scripts like *Śāradā* and *Grantha*.

AI and Machine Learning Applications

ML and AI may be used to enhance accuracy and efficiency in manuscript decoding.

Handling Complex Manuscript Structures

Future technologies must faithfully capture illustrations, marginal notes, and layered annotations.

Advancement of Digital Tools

Specialized tools for transcription, transliteration, annotation, and preservation should be developed and integrated.

Decoding Manuscripts: Standards to be Followed

The process of decoding manuscripts demands adherence to international and national standards to ensure accuracy, interoperability, and long-term usability of the transcribed and digitized content. Standards such as Unicode provide a universal character encoding system that enables the representation of diverse scripts, diacritical marks, and special symbols found in Indian manuscripts, thereby making texts accessible across various platforms and technologies. The Text Encoding Initiative (TEI) provides a comprehensive framework for representing the structural, semantic, and interpretive features of texts, enabling the detailed markup of variations, corrections, commentaries, and annotations essential in manuscript studies. Similarly, the International Image Interoperability Framework (IIIF) standards facilitate the sharing, comparison, and collaborative research of high-resolution images, while OCR/HTR standards (Optical Character Recognition/Handwritten Text Recognition) align with machine-learning models for script-specific decoding. Following these standards ensures that transcription, encoding, and digital preservation of manuscripts not only remain faithful to the original but also integrate seamlessly into global digital knowledge networks. Establishing such consistency is essential for scholarly collaboration, sustainability, and future-proofing the manuscript heritage.

Insights from IKS for Textual Interpretation

Bharata has a rich linguistic tradition, and the principles of the IKS help provide valuable insights for textual interpretation. They highlight how indigenous methods of understanding words and meanings, structure, and context can illuminate the decoding of manuscripts and scripts.

The Central Role of Hermeneutics in Decoding Scripts and Manuscripts

“lipidarśanam śabdasmṛtiṃ janayitvā śābdabodhaṃ prayojayātīti”

(Tarkasaṅgraha-bālapriyā-vyākhyā by Śrīrāmānujātātācārya)

This statement mentions that the mere sight of written letters (*lipidarśana*) generates remembrance of words (*śabdasmṛti*), which in turn produces verbal cognition (*śābdabodha*). It emphasizes that writing is not an end in

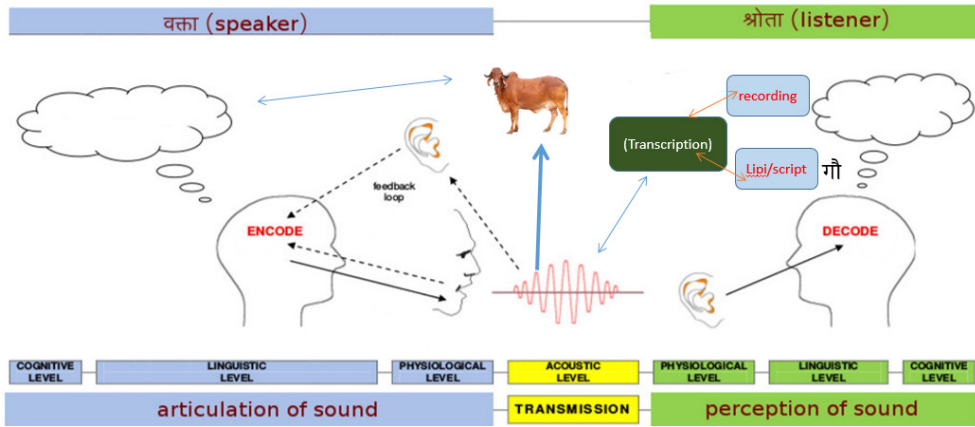
itself but a means to meaning, whether through oral expression or silent reading²². *Śābdabodha* is therefore central to the decoding process, as it connects visible symbols to intelligible knowledge. In manuscript traditions, accurate recognition of letters only gains value when it leads to correct comprehension. This underscores the role of grammar, oral tradition, and memory in interpreting texts. Without *śābdabodha*, even perfectly preserved scripts remain lifeless. Thus, for both traditional scholarship and modern technologies like OCR and transliteration, the ultimate aim must be enabling *śābdabodha* — the actual understanding of meaning from the script.²³

Again, through different *pramāṇas* (means of valid knowledge), a person acquires understanding of reality. However, this knowledge remains internal until it is expressed. To share it with others, the knower encodes it into language, shaping thought into words and script. For the hearer or reader, mere exposure to words is not enough—the decoding process must culminate in *śābdabodha*, where the intended meaning is grasped as a unified whole. Thus, *pramāṇa* provides the foundation of knowledge, language provides the medium of its transmission, and *śābdabodha* ensures its proper reception.

The above-mentioned points are summarised in the following diagram-

22 When a person reads silently, the eyes perceive written symbols (*lipidarśana*), which are instantly mapped onto their corresponding sounds and meanings in the mind. Though there is no outward voice, the brain simulates inner speech—engaging the same cognitive pathways as oral recitation. This inner articulation activates memory, imagination, and comprehension, allowing the reader to grasp concepts, relate them to prior knowledge (*śābdasmṛti*), and form new insights. Silent reading thus transforms visual signs into mental sound and sense, enabling knowledge acquisition (*śābdabodha*) without audible expression. In this way, script empowers both external communication through recitation and internal cognition through silent engagement.

23 Also see the *Sopajñāvr̥tti*, where it is stated: “*padajñānaṃ tu karaṇaṃ dvāraṃ tatra padārthadhīḥ, śābdabodhaphalaṃ tatra śaktidhīḥ saha cārīṇī*.” This means that knowledge of words serves only as an instrument and gateway to their referents, while the true outcome lies in *śābdabodha*—the comprehension of sentence-meaning accompanied by awareness of its semantic power. In the context of manuscript studies, this indicates that mere identification of words or letters is insufficient; authentic interpretation requires moving beyond word-level recognition to grasp the intended meaning.



The Concept of Text and Manuscript in the *Jāti–Vyakti* Framework

The distinction between text and manuscript lies at the heart of textual interpretation. This can be expressed in terms of *jāti* (type) and *vyakti* (token) distinction found in the *śāstras* of IKS. While the text represents the intellectual composition intended by the author, the manuscript is its material witness, shaped by historical processes of transmission and preservation. Philosophical categories such as *jāti* and *vyakti* provide a refined lens for understanding how multiple manuscript copies, which can be called *vyakti*, embody but never fully exhaust the underlying text, which can be viewed as *jāti*. Recognizing this distinction allows scholars to approach manuscripts not as final authorities but as gateways to reconstructing the archetypal composition.

4.2.1 Concept of Text and a Manuscript

In textual criticism, a fundamental distinction is made between the *text* and the *manuscript*. The *text* refers to the intellectual content or composition as envisaged by the original author, while the *manuscript* is a particular material instantiation of that text, copied, transmitted, and preserved in palm leaves, birch bark, paper, or digital form. A manuscript may faithfully preserve, partially distort, or substantially alter the author's words through scribal errors, interpolations, omissions, or regionally conditioned variations. Hence, decoding a manuscript requires disentangling the author's intended composition from its specific written embodiment. While manuscripts are the only vehicles through which ancient written texts survive, they are not the texts themselves, but witnesses to them.

4.2.2 Problem of *Jāti* and *Vyakti* (Type and Token)

Bhāratīya philosophical categories of *jāti* (universal/type) and *vyakti* (particular/token) illuminate the challenge of textual criticism. The *text* may be conceived as a *jāti*—a universal entity that transcends any one manuscript—while every manuscript copy is a *vyakti*, a token instantiation of that universal. No single manuscript may perfectly embody the text, but each provides a partial and historically conditioned access to it. Variants among manuscripts thus represent different *vyaktis* of the same *jāti*. The task of the critic is to compare, classify, and collate these *vyaktis* in order to approximate the underlying *jāti*, the archetypal text intended by the author. This theoretical framing clarifies why textual criticism does not treat manuscripts as ends in themselves, but as evidence to reconstruct the type that they imperfectly manifest.

4.2.3 Properties of Text and Manuscripts

Texts and manuscripts differ in their properties. The *text* is abstract, stable in principle, and tied to authorial intention. It is an intellectual construct that does not perish with the loss of a single manuscript but remains potentially retrievable through other copies or traditions. The *manuscript*, by contrast, is concrete, material, and historically situated, subject to wear, loss, scribal intervention, and linguistic evolution. Its properties include orthography, scribal notations, marginalia, post-colophons, and local linguistic idioms—all of which may illuminate the context of transmission but can also obscure the authorial text. Thus, the textual editor must weigh both the reliability of the manuscript witness and the nature of textual variation. Through processes like recension, stemmatic analysis, and emendation, textual criticism attempts to move from the multiplicity of *vyaktis* to the unity of the *jāti*, producing a critical or comprehensive edition. Critical Edition is that which approximates, as closely as possible, the author's original composition.

4.2.4 Problem of Published or Unpublished

In the field of manuscript studies and textual criticism, the distinction between “published” and “unpublished” is not as straightforward as it appears. A *text* as an intellectual entity (*jāti*) may be well known, cited, and commented upon, yet its surviving *manuscripts* (*vyakti*) may remain unpublished or only partially edited. Conversely, a single manuscript may be printed in a facsimile edition, yet that does not guarantee that the *text* it embodies has been critically/comprehensively established or truly “published” in the scholarly sense. The act of publication, therefore, lies not merely in printing or digital reproduction

but in the scholarly process of transcription, transliteration, collation, and critical editing that aims to approximate the author's original composition. Without such critical engagement, even a "published" manuscript may still conceal more than it reveals.

1.1.5. Who Decides if a Manuscript is Published?

The question of who decides whether a manuscript is considered published reflects a deeper methodological problem. Publishers, libraries, or digital archives may classify a manuscript as "published" once it is made available in print or online, but from the perspective of textual criticism, such a label is provisional. For scholars, true publication is achieved only when the *text* has been critically reconstructed through the comparison of multiple manuscripts, the correction of errors, and the contextual interpretation. Thus, while the availability of a manuscript reproduction may be decided by institutions, its status as a "published text" must ultimately be determined by the community of scholars who evaluate whether the critical apparatus adequately represents the underlying *jāti* rather than a single *vyakti*.

4.2.6 *Prākṛtadhvani* and *Vaiṣṇavadhvani* in Textual Transmission

When a scribe writes a manuscript, there is a component of attribute that distinguishes one manuscript from another, such as handwriting style, orthography, or regional idiom. In textual criticism, these attributes must be carefully separated to approximate the author's intended text.²⁴ This can be expressed in the terminology of IKS through the distinction made between *prākṛtadhvani* and *vaiṣṇavadhvani*. *Vaiṣṇavā*, when articulated by a person, has a component of *prākṛtadhvani*, representing the natural quality of sound, and *vaiṣṇavadhvani*, representing incidental distortion.

4.2.7 Manuscript Integration Platforms

A recurring problem is the fragmentation of manuscript traditions. A text may survive in dozens of manuscripts scattered across libraries, temples, or private collections. Without integration, scholars often work in isolation on partial evidence, producing incomplete or conflicting editions. Digital platforms that unite scattered holdings can address this problem by enabling comparison across manuscripts, supporting cross-referencing of variants, and

24 The same principle applies in modern domains: speech recognition systems filter natural variations and distortions to capture intended words, while OCR must distinguish script features from errors caused by degradation or noise.

reconstructing complete transmission histories. Such integration helps resolve the *jāti*–*vyakti* tension: the platform enables the collation of many *vyaktis* in pursuit of the underlying *jāti*.

Semantic Keys from *Śāstra* for Interpreting Manuscripts

Traditional *śāstras*—particularly *Vyākaraṇa*, *Mīmāṃsā*, and *Alaṅkāraśāstra*—provide a sophisticated semantic framework for decoding manuscripts.

The principles of *saṃyoga* (conjunction) and *viprayoga* (separation) highlight how the joining or splitting of letters and words—whether through *sandhi* or ligatures—can alter meaning, and in manuscripts, any misjoining or break in script may distort sense. Similarly, *sāhacarya* (association) and *virodhitā* (opposition) remind us that the meaning of a passage depends on the compatibility or contrast of neighboring terms, making contextual co-occurrence crucial in decoding. Broader factors such as *artha* (sense), *prakaraṇa* (context), *liṅga* (semantic indicators), and *śabdasya anyasya saṃnidhiḥ* (the presence of other words) serve as interpretive guides whenever expressions are ambiguous or polysemous. To ensure coherence, *sāmarthyā* (syntactic and semantic fitness) must be respected, while *aucitī* (propriety or aptness) ensures that meaning aligns with cultural and stylistic appropriateness. Moreover, context is shaped by *deśa* (place), *kāla* (time), and *vyakti* (the speaker or subject), all of which influence the nuance of expression. Finally, *svarādayaḥ* (intonation, accents, and phonetic features) are especially significant in Vedic and poetic texts, where pitch and sound modulations affect meaning. Taken together, these categories constitute a hermeneutic toolkit—semantic keys that guide the scholar in decoding manuscripts faithfully beyond mere transcription.²⁵ By integrating these interpretive tools, scholars can move beyond mechanical transcription to authentic understanding, ensuring that the cultural and intellectual depth of the manuscript tradition is preserved in its decoded form.

In *Nyāyasiddhāntamuktāvalī*, it is explained that the grasp of meaning (*śaktigraha*) does not arise in a vacuum but through multiple sources of semantic orientation. The elders (*vrddhāḥ*), i.e. authoritative teachers, state that it comes from grammar (*vyākaraṇa*), analogy (*upamāna*), lexicons (*kośa*), reliable statements (*āptavākya*), and also from common worldly usage (*vyavahāra*). Furthermore, comprehension develops from the remainder of a sentence (*vākyasya śeṣa*), from explanatory commentary or

25 *saṃyogo viprayogaś ca sāhacaryaṃ virodhitā / arthaḥ prakaraṇam liṅgam śabdasya anyasya saṃnidhiḥ // sāmarthyam aucitī deśaḥ kālo vyaktiḥ svarādayaḥ / śabdārthasya anavacchede viśeṣasmṛtihetavaḥ//*

clarification (*vivṛti*), and from the contextual proximity of established words (*sānnidhyataḥ siddhapadasya*). In the context of decoding manuscripts, this principle underscores that understanding a doubtful or obscure word depends on cross-verification through grammar, comparison, dictionary support, trusted textual tradition, and practical usage. When words are damaged, rare, or contextually ambiguous, meaning must be reconstructed by attending to sentence completion, scholastic explanations, and the association with already intelligible words.²⁶

The Role of Prakaraṇa (Context)

In the process of context-based textual determination, not only knowledge of individual letters but also familiarity with the technical terminology of disciplines such as astrology and medicine, metrical patterns, numerical systems, and the principles of *sandhi* and *samāsa* is essential. At times, a word or phrase may admit more than one reading, both of which are grammatically and metrically valid, yet yield very different meanings. For example,

1. *mūrkhahaste na dātavyam* (“knowledge should not be given into the hands of a fool”) and
2. *mūrkhahastena dātavyam* (“it should be given by the hand of a fool”)

Here, both transcriptions are linguistically correct, but the intended sense is determined only by the context. For traditional scholars, such cases did not pose a difficulty, as the surrounding subject matter guided interpretation. However, for modern transcription into standardized scripts—where the concept of spacing and word division was not always clear in the original manuscripts—these variations may create puzzles.

In manuscript decoding, knowledge of *śabda-artha* (semantic awareness) is as essential as script literacy. Words formed through *sandhi*, *samāsa* (compounds), *bhūtsaṅkhyā* encoding (word-number system), and other means often present multiple interpretative possibilities. If the scribe relies only on the visible spacing or breaks between words, unintended meanings may arise. Hence, the transcription must be guided by contextual appropriateness rather than surface segmentation. Similarly, a correct understanding of *chandas* and technical terminology is vital for establishing the *śuddhapāṭha* (accurate reading). Where variant readings exist, the one consistent with the overall

26 *śaktigrahaṃ vyākaraṇopamānaśāptavākyād vyavahārataś ca/ vākyasya śeṣād vivṛter vadanti sānnikidhyataḥ siddhapadasya vṛddhāḥ//* (Nyāyasiddhāntamuktāvalī)

prakaraṇa (subject matter) should be preferred.

Use of Bharatiya Traditions for Textual Criticism

The manuscripts embody the living tradition of knowledge, enriched through commentaries, marginal notes, scholastic refinements, and regional variations, showing how ideas were continuously debated, tested, and reshaped across generations. Their distinctive knowledge organization—through *bhāṣya*, *vyākhyā*, *ṭīkā*, *ṭippaṇa*, *ṭupṭīkā*, *pañjikā*, and further sub-commentaries—along with translations into diverse languages, stands as an unparalleled intellectual tradition in human history. This unique tradition aids in manuscript research by identifying the version of the text that was available to the commentator at the time the commentaries or regional language translation was composed.

Enhancing Textual Clarity with Traditional and Modern Marks

The use of modern symbols and concepts has played a crucial role in the standardization of language, particularly through the influence of scripts. Different marks used in writing make meanings clearer. In various scripts, especially in Vedic texts, specific marks were used to indicate vowels, such as *udātta*, *svarita*, *dīrghasvarita*, *jihvāmūlīya*, *upadhmānīya*, *nāsikya*, as well as other branch-specific signs. Punctuation marks such as the *virāma* [ı] (full stop) and the *yugala-virāma* [ıı] (double-full stop for the end of a *śloka*) were also naturally found in these scripts. These marks help in understanding sentence divisions, thematic divisions, the completion of verses etc. In some scripts, rare marks such as the *avagraha* [ɤ] and other *śāstra*-specific markers were also available—for example, special notations in *Saṅgītaśāstra* (musical symbols for *svara* and *tāla*), in *Jyotiṣa* (astronomical and astrological symbols for planets and zodiac signs), and in *Gaṇita* (mathematical symbols for numbers, operations, and fractions).

In addition to the symbols found in a manuscript, modern punctuation marks become necessary in transliteration and transcription to make texts more legible and human-readable, clarify sentences, define grammatical structures, divide or emphasize meanings, and indicate relationships between ideas. These include Comma “,” (*alpavirāma*), Semicolon “;” (*ardhavirāma*), Colon “:” (*dviprāṇa*), Question Mark “?” (*praśnavācaka-cihna*), Exclamation Mark “!” (*vismayādibodhaka* / *āścarya-cihna*), Apostrophe “'” (*uddharana*), Quotation Marks “ ”” (*uddharaṇa cihna*), Parentheses “()” (*koṣṭhaka*), Square Brackets “[]” (*samacatura-koṣṭhaka*), Ellipsis “...” (*virāmabindu*), Hyphen “-” (*vojaka cihna*), Slash “/” (*tiryak cihna*), Underscore “_” (*adhorekhā*), Caret “^” (*pāṭha sampādana prayukta*), Asterisk “*” (*tāraka*) etc.

Along with punctuation and other marks, concepts such as space “ “ (*avakāśa*) are also indispensable for accurate transliteration and reading of texts. The concept of *avakāśa* is rare in traditional manuscripts, where words were often written continuously without clear separation. The absence of proper spacing creates significant challenges in transliteration, as illustrated by the following example.

*śāradāṃ varadāṃ vande prasannavadānāṃ śubhāṃ/
vidyāṃ buddhiṃ tathā jñānaṃ natebhyo dehi vatsale//*

(I bow to *Śārada*, O dear one, the gracious giver of boons, whose face is serene and auspicious; bestow knowledge, intellect, and wisdom upon those who bow before You.)

Here, the last *pāda* of this verse “*na tebhyo dehi vatsale*” can be transcribed in multiple ways:

1. *natebhyodehivatsale* – (a continuous script as seen in the manuscript without any spaces, making it difficult to parse the individual words.)
2. *na tebhyo dehi vatsale* – (a possible segmented reading, meaning ‘O dear one, do not give to them,’ where ‘*na*’ functions as a negation, though unintended in this context.)
3. *natebhyo dehi vatsale* – a correct segmentation, meaning “Give to the bowed ones, O dear one,” where ‘*nata*’ is read as “bowed”, demonstrating a semantic shift caused by the absence of space.

These examples illustrate how spacing (*avakāśa*) in transliteration directly affects grammatical parsing and semantic interpretation. Manuscripts often lack clear spaces, so proper segmentation is essential to preserve meaning. Even a small space—or its absence—can completely alter the reading, highlighting the critical role of *avakāśa* in modern textual transcription.

Manuscripts and the Living Tradition of IKS

Manuscripts serve as vital pathways to the IKS because they preserve, transmit, and contextualize the intellectual, spiritual, and cultural traditions

of Bharata across millennia. They hold within them the accumulated wisdom of diverse fields, providing not only the content of knowledge but also insights into the methods of learning, interpretation, and dissemination. As physical repositories of knowledge, manuscripts reveal the material culture of writing—scripts, calligraphy, ornamentation, and preservation techniques—opening windows into the social, cultural, and institutional practices of their times.

Conclusion

Each manuscript holds within it the echoes of oral traditions, the precision of grammar, and the depth of philosophical insight, encoded through intricate scripts and preserved across centuries. The process of decoding is, therefore, both a scholarly pursuit and a civilizational responsibility—one that transforms dormant texts into living wisdom, relevant to contemporary thought and future innovation.

“bhūyāt bhāratapañkajaṃ kalimāla-pradhvaṃsi naḥ śreyase”

(May the *bhāratapañkaja*, which destroys the impurities of the Kali age, flourish once again for our well-being (or ultimate good).)

(Just as a lotus emerges unsullied from the mud, Bharata shall rise again—rooted in its ancient knowledge and radiant with wisdom for the world. The very name *Bhārata* (*bhāsi ratam prāpnoti iti bhāratam*) implies “that which is devoted to brilliance or knowledge.” In this light, “ज्ञानभारतम्” symbolizes the reawakening of Bharata as *Viśvaguru*—a guiding light for the world, not through dominance, but through the shared wealth of its timeless intellectual and spiritual traditions. Manuscripts in philosophy, science, art etc.—once the lifeblood of Bharata’s flourishing—now serve as pathways for its resurgence through knowledge.)

References

- Allen, W. S. 1953. *Phonetics in Ancient India*. Oxford: Oxford University Press.
- Bhat, Vasantakumar. *Pāṇḍulipi evaṃ Samīkṣit Pāṭha-Sampādana*. Indira Gandhi National Centre for the Arts, New Delhi, 2025, p. 6.

- Bhāṭiyā, Omprakāśa. 2019. *Lipivijñāna aura Nāgarī Lipi*. Dillī: Sūrya Prakāśana, Naī Saṛaka.
- G. H. Bhatt (Ed.). *The Vālmīki Rāmāyaṇa: Critical Edition*. Baroda: Oriental Institute, 1960–1975.
- Gode, P. K. (1953–1960). *Studies in Indian Cultural History* (Vols. 1–2). Poona.
- Jagannātha, S. 2007. *Manuscriptology: An Entrance*. Delhi: Parimal Publications.
- Kulkarnī, Ambā. *Devanāgarī Lipi aura Saṅgaṇaka*.
- Ojha, Gaurīśaṅkara Hīrācanda. 1894. *Bhāratīya-Prācīna-Lipimālā* (The Paleography of India). Naī Dillī: Munshī Rām Manoharlāl.
- Pāṇḍe, Rājabalī. 2018. *Bhāratīya Puralipi*. Ilāhābād: Lokabhārati Prakāśana.
- S. Jagannatha. *Manuscriptology: An Entrance*. Delhi: Parimal Publications, 2007. ISBN 9788171103157.
- S. Jagannatha. ग्रन्थसम्पादनशास्त्रम् (*Granthasampādanaśāstram*). बेङ्गलूरु: पूर्णप्रज्ञसंशोधनमन्दिरम्, 2014. ISBN 9789381275399.
- S. M. Katre. *Introduction to Indian Textual Criticism*. Poona: Deccan College, 1941 (later reprints).
- Salomon, Richard. 1998. *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan Languages*. New York: Oxford University Press.
- Satyendra. *Pandulipi Vijñāna*. Jaipur: Rajasthani Granthagar, 1978. (Reprints available).
- Scharf, Peter & Malcolm Hyman (Author). *Linguistic Issues in Encoding Sanskrit*.
- Shamashastri, R. (1909). *The Kautilīya Arthaśāstra*. Mysore: Government Oriental Library Series.
- Shivaganesh Murthy, R. S. 1996. *Introduction to Manuscriptology*. Cambridge: Cambridge University Press.
- Upasak, C. S. 2001. *History and Palaeography of Kharoṣṭhī Script*. Sarnath: Central Institute of Higher Tibetan Studies
- V. N. Jha (Ed.). *Problems of Editing Ancient Texts*. Delhi: Chaukhamba

Sanskrit Pratishthan, 1993.

- V. S. Sukthankar (Gen. Ed.). *The Mahābhārata: Critically Edited*. Poona: Bhandarkar Oriental Research Institute, 1927–1966. *Prolegomena* (Vol. I, Ādiparvan) first published 1933.
- V. S. Sukthankar. *Prolegomena to the Critical Edition of the Mahābhārata (Ādiparvan)*. Poona: Bhandarkar Oriental Research Institute, 1933.



Working Group VII

Theme: Manuscripts as Tools of Cultural Diplomacy

Coordinator

Prof. Anirban Dash

Director, National Mission for Manuscripts (NMM)

Co-Coordinator

Shri Aneish P. Rajan

Director, Ministry of Culture, Govt. of India

Experts

Dr. V. Ramanathan

Associate Professor, Dept. of Chemistry, IIT BHU

Prof. Venketeswara Pai R.

Associate Professor, Dept. of Humanities & Social Sciences, IISER Pune

Dr. Amogh Dev Rai

Executive Director, Sanrachna, SGT University

Prof. Siniruddha Dash

Former Prof. and Head, Chief Editor, New. Cat. Catalogorum (NCC),
Department of Sanskrit, University of Madras

Prof. Chander Shekhar

Professor, Dept. of Persian, University of Delhi

Shri Abhay Kumar

Deputy Director General (C&SP), ICCR

Manuscript as a Tool to Cultural Diplomacy

Manuscripts stand as timeless witnesses to India's civilizational journey, inscribed in diverse Languages, Scripts, and disciplines. They embody not just the written word but the very essence of India's cultural and intellectual wealth. Manuscript knowledge has traveled across regions and generations in two interwoven forms: as lipi-bhāṣā setu—the bridge of scripts and translations —and as jñāna-dharohara, the treasure of content, such as the Indian epic Ramayana and the Buddha Dhamma, which have shaped global thought, culture, and practice.

With millions of manuscripts scattered across the globe in temples, mutts, libraries, and private collections, India safeguards the world's largest reservoir of written wisdom spanning Philosophy, Science, Medicine, Mathematics, Art, and Spirituality.

By weaving the manuscript tradition into international engagements, India can foster cultural affinity and mutual respect, building connections that transcend borders and ideologies. In this way, Manuscripts become more than repositories of knowledge; they become instruments of global harmony, carrying forward India's eternal message of *Vasudhaiva Kuṭumbakam* - “the world is one family.”

1. India's Place in the Global Knowledge Tradition

For centuries, Manuscripts have been India's cultural ambassadors, translated and transmitted across cultures, and have long served as cultural bridges. The origin of diverse languages and scripts across Asia, for instance, Sinhalese in Sri Lanka, Tibetan in the Himalayas, and various other scripts that influenced the world, reflects India's profound *Bhartiya gyan* contribution, demonstrating how its intellectual and literary traditions shaped civilizations far beyond its borders.

Sanskrit treatises on Mathematics and Astronomy influenced developments in algebra and trigonometry across Europe, sometimes directly through translated manuscripts. British and German universities, in particular, built “*Orientalist*” scholarship and even their reputational foundations based on the study of Indian manuscripts such as the *Upaniśads*, *sūtras*, and scientific works.

The *Rigveda*, translated into English by Max Müller, opened the wisdom of the Vedic hymns to the Western world, laying the foundation for comparative religion and Indology. Similarly, *the Pañcatantra*, *Sukasapatati*, and *Kālidāsa's Abhijñānaśākuntalam*, among many others, have been rendered into various

foreign languages, stirring admiration in literary circles worldwide and influencing thinkers, scholars, poets, philosophers, and travelers. Such works serve as a testament to the influence of Indian manuscripts on global literary and intellectual discourse.

Yet, the dispersal of manuscripts tells a complex story. Many Indian manuscripts have been traced abroad- some through scholarly exchange and cultural curiosity, others illicitly transferred, sold, or smuggled. Yet wherever they reside, they continue to act as silent ambassadors of India's knowledge systems, influencing societies far beyond their place of origin.

2. Institutionalize Manuscript-Centric Cultural Diplomacy

Manuscripts need to be recognized as strategic instruments of soft power within India's cultural diplomacy framework. Currently, manuscript-related initiatives remain fragmented, i.e., they are limited to occasional exhibitions, academic conferences, or individual collaborations. To unlock their full potential, a dedicated national programme should be established under the Ministry of External Affairs in partnership with the Ministry of Culture, ICCR, and the National Mission for Manuscripts.

Such a programme can:

- Integrate manuscripts into bilateral and multilateral cultural agreements.
- Establish manuscript exchange fellowships for Scholars and Conservators.
- Create cultural centres abroad that display Manuscripts as symbols of India's knowledge tradition.
- Leverage digitization projects for global visibility while safeguarding fragile originals.
- This Institutionalization ensures continuity, coordination, and a clear roadmap, transforming manuscripts from passive heritage into active agents of international engagement.

3. Global Exhibitions and Cultural Showcases

International exhibitions can act as cultural windows into India's intellectual legacy. Manuscripts curated thematically—on Ayurveda, Astronomy, Buddhist Philosophy and Psychology, Mathematics, classical literature, and

many more — can speak directly to contemporary global concerns, such as health, sustainability, peace, and ethics.

- **Partnerships with Global Institutions:** Collaborations with the British Library, Library of Congress, and Asian manuscript repositories can make India's collections accessible worldwide.
- **Traveling Exhibitions:** Manuscripts showcased in major world capitals would deepen cultural ties and showcase India's civilizational depth.
- **Digital Showcases:** High-quality digital curation and immersive technologies exhibitions would democratize access, reaching audiences who cannot visit physical displays.

By aligning these initiatives with forums like G20 cultural platforms, India positions its manuscript heritage not only as a national treasure but as a shared global legacy, reinforcing its role as a thought leader in culture, knowledge, and diplomacy.

4. **Facilitate Global Academic and Research Collaboration**

Manuscripts thrive not just in preservation but in reinterpretation and renewal. By fostering global academic and creative collaborations, India can ensure that its manuscripts remain relevant to the evolving intellectual and cultural landscape.

- **Joint Research and Translation Projects:** Scholars across the globe can collaborate on editing, diplomatic and critical editions, translations into other scripts and languages, and contextualizing manuscripts, making ancient knowledge accessible in contemporary languages.
- **Creative Reinterpretations:** Manuscripts can inspire theatre productions, films, digital storytelling, and artistic projects that resonate with modern audiences.

For example, epics and dramas preserved in manuscripts can be reimaged for global cinema or theatre.

- **Academic Networks:** Collaborating with Universities, Repositories, Schools and establishing International Manuscript Chairs, Summer Schools, and Exchange Programmes can position India as the hub for Manuscript Studies.

5. **Launch International Fellowships and Scholarships**

Fellowships and scholarships dedicated to manuscript studies can serve as powerful instruments of cultural diplomacy by drawing global scholars, psychologists, linguists, and artists into India's intellectual orbit.

- **Global Engagement:** Fellowships invite and encourage National and International Scholars, Psychologists, Linguists, Artists, and those from various other disciplines to study India's Manuscript Heritage, creating direct avenues of cultural and intellectual exchange.
- **Contemporary Relevance:** Research themes can be designed around issues of global concern. Such as mental well-being, ethics, sustainability, and intercultural harmony, demonstrating how ancient wisdom addresses modern challenges.
- **Decoding and Translation:** Projects can focus on decoding manuscripts and translating them into Indian languages and scripts, making knowledge accessible to wider audiences while preserving linguistic diversity.
- **Short- and Long-Term Projects:** Both quick-impact and in-depth research projects can be supported, generating international recognition for India's manuscripts and sustaining scholarly engagement over time.
- **Fellowship Networks:** Building global fellowship ties, including visiting fellowship programs, creates lasting academic and cultural bridges. Scholars return as cultural ambassadors, carrying India's wisdom traditions back to their own societies.
- **Soft Power Diplomacy:** These initiatives foster goodwill, respect, and long-term intellectual partnerships, positioning India as a hub of manuscript studies and a global center of holistic knowledge and civilizational dialogue.

6. Declare and Promote a Global 'India Manuscript Day'

- Celebrating manuscripts on Guru Purnima Day symbolizes India's rich cultural and intellectual heritage.
- Spread global awareness about the value of manuscripts across disciplines and traditions.
- Inspire greater participation of scholars, students, and cultural

enthusiasts in manuscript studies.

- Encourage initiatives to decode, translate, and interpret manuscripts for contemporary relevance.
- Build international dialogue and collaboration, positioning manuscripts as bridges of shared human wisdom.
- Strengthen cultural diplomacy by showcasing India's role as a custodian of civilizational knowledge.

7. Leverage Manuscripts in Digital Cultural Diplomacy

By digitizing, curating, and creatively presenting manuscripts, India can extend its knowledge traditions beyond archives and reach a truly global audience.

- **Virtual Exhibitions:** Manuscripts can be showcased online with immersive formats featuring 3D archives, virtual reality tours, and interactive storytelling will bring ancient manuscripts to life for global reach.
- **Accessibility to Multilingual Translations:** Making manuscripts accessible in multiple world languages, as well as decoding them into Indian languages and scripts under one umbrella, i.e., National Digital Repository (NDR), fosters inclusivity and broadens participation.
- **Digital Storytelling:** Integrating insights from traditional knowledge and its practices, such as script learning and writing, into apps, podcasts, and online learning platforms connects ancient wisdom with contemporary psychological and ethical debates.
- **Youth Engagement:** By leveraging AI, gamification, and digital media, manuscripts can capture the imagination of younger generations, inspiring curiosity and pride in India's civilizational legacy.

8. Advance Repatriation and Legal Protection of Manuscripts

Many invaluable manuscripts lie scattered in foreign archives, private

collections, and institutions. Advancing their repatriation through cultural diplomacy, international cooperation, and legal frameworks strengthens India's role as a responsible custodian of civilizational wisdom.

- **Repatriation Diplomacy:** Engage in dialogue with countries and institutions to bring home manuscripts of Indian origin, treating them as cultural treasures and shared heritage.
- **Legal Mechanisms:** Strengthen national and international legal mechanisms to prevent illicit trade, smuggling, or misappropriation of manuscripts.
- **Ethical Collaboration:** Promote joint custodianship models, digitization partnerships, and knowledge-sharing agreements where physical return is not possible.
- **Global Awareness:** Use forums like UNESCO and cultural diplomacy platforms to highlight the importance of manuscripts as living embodiments of human heritage.

9. Develop Thematic Manuscript Heritage Circuits

India's ancient centres of learning are living monuments to its civilizational brilliance. By weaving manuscript repositories into manuscript tours, these sites become gateways for both domestic and international visitors to experience India's intellectual and spiritual heritage.

- **Highlighting Ancient Universities and Monastic Centres:** Establish circuits connecting Nalanda, Vikramshila, Odantapuri, Sringeri, Thanjavur, Varanasi, Kashmir, and other hubs of scholarship, emphasizing their historical role in knowledge creation, preservation, and transmission.
- **Integration with Manuscript Repositories:** Visitors engage directly with digitized and physical manuscripts, exploring manuscripts in Sanskrit, Pali, Tamil, Grantha, and regional scripts, bringing alive the Pan-Indian knowledge continuum.
- **Regional cooperation:** Multilateral efforts with BIMSTEC, SAARC, and ASEAN nations emphasize language exchanges, heritage studies, and the creation of regional research consortia.
- **Thematic Routes:** Manuscript museum on themes such as Buddhist Philosophy, Classical Arts, Science and Mathematics, Ayurveda, and

Ethics, connecting manuscripts with tangible cultural sites, temples, libraries, and museums.

10. Thematic Manuscript Festivals

- **Experiential Engagement:** Visitors explore manuscripts through interactive workshops, storytelling sessions, live readings, calligraphy demonstrations, and traditional performance arts, bringing the manuscripts to life. For instance, Holi celebrations, as illustrated in manuscripts, can be recreated through live performances and Madhubani painting workshops, blending color, tradition, and literary heritage.
- **Celebrating Women:** Manuscripts often reflect the role, wisdom, and accomplishments of women in various spheres. Women's Day-themed events can highlight these references through readings, discussions, and exhibitions, celebrating the voices of women preserved in India's literary heritage.
- **Celebrative Collaboration:** Schools, universities, research institutes, and local communities participate, fostering **intergenerational knowledge transfer** and cultivating **pride in India's intellectual legacy**.

11. Expand Educational Diplomacy through Indic Knowledge Systems

- **Curricular Integration:** Universities and Schools worldwide incorporate Manuscripts into multidisciplinary curricula, covering Ayurveda, Yoga, Buddhist Psychology, Classical Literature, Astronomy, Mathematics, and the Arts.
- **Align with New Education Policy 2020** and allied policies encourage multidisciplinary universities, flexible learning paths, and experiential education rooted in Indic Knowledge Systems (IKS)
- **Scholarships and Fellowships:** India launches international fellowships and research grants to study manuscripts as streams in

schools/colleges, inviting scholars, students, and practitioners globally to engage with manuscripts, critical editions, and digitized archives.

- **Cross-Border Research Collaboration:** Joint Programs with Institutions in Europe, Asia, and the Americas encourage comparative studies, interdisciplinary projects, and innovations inspired by Indian knowledge systems, fostering global academic networks.

These networks would not only disseminate knowledge but also integrate the Scientific and holistic dimensions of Indian thought systems. For instance, mind management in Buddhism (*citta-bhāvanā*) and the balance of the three guṇas in Sāṃkhya are frameworks for addressing modern challenges, such as stress, anxiety, and societal disharmony.

12. Manuscripts as Living Knowledge: Scientific and Contemporary Relevance

These manuscripts are intellectual reservoirs that hold the potential to enrich Contemporary Science, Ecological discourse, Healthcare Paradigms, Ethical Frameworks, and Interdisciplinary Scholarship.

Consider a few instances:

- ***Bhūkampa Lakṣaṇa:*** An early Indic treatise detailing the causes, precursors, and cyclical patterns of earthquakes, reflecting an indigenous understanding of natural disasters.
- ***Gajakitsā:*** A text on veterinary medicine, particularly the diagnosis and treatment of elephants, showcasing ancient India's expertise in zoology and animal husbandry.
- ***Kṛṣiparisara:*** Manuscripts offering insights into sustainable agricultural practices, soil management, seasonal cycles, and crop rotation.
- ***Khagolasāra:*** A brilliant exposition on astronomical instruments and celestial measurements, demonstrating the precision and observational acumen of Indian Astronomers.
- ***Prthvī Lakṣaṇa:*** A remarkable treatise detailing the characteristics, qualities, and energies of the Earth. This manuscript explores terrestrial signs, geographical classifications, land fertility, and natural phenomena, offering indigenous models of earth science, geophysics, and ecological consciousness.
- ***Brahmāṇḍabhūgola:*** A cosmological treatise integrating terrestrial

geography with celestial mappings, reflecting a holistic Indic understanding of the universe, Earth's position within it, and the interrelation between spatial, spiritual, and natural orders.

In an era marked by complex challenges ranging from environmental sustainability and health crises to ethical dilemmas and technological advancements, these manuscripts may offer critical perspectives that can contribute to holistic and grounded solutions. Their study is not merely an academic pursuit but a pressing need to revive and integrate India's indigenous knowledge systems into modern frameworks.

13. Multinational Manuscript Digitization Projects

- **Global Collaboration:** Partner with international libraries, universities, and research institutions to digitize Indian manuscripts.
- **Accessibility:** Make manuscripts available worldwide for scholars, students, and the general public through online platforms.
- **Subscription Models:** Offer access to digitized manuscripts through institutional and individual subscriptions for universities, research centers, and libraries globally.
- **Collaborative Digitization Grants:** Partner with international institutions on funded projects using digitized manuscripts, attracting research grants.
- **Publication Rights and Licensing:** Publish annotated translations, critical editions, or multimedia content for commercial distribution.
- **Manuscript Tour Tie-ins:** Link digitized content with thematic manuscript circuits or heritage tours, promoting paid cultural tourism packages.

14. Manuscripts as Soft Power in Geopolitical Arenas

India's manuscript heritage can be a powerful tool in diplomatic relations, particularly in regions with historical or cultural ties to the country.

- **Strengthening ties with Southeast Asia:** Manuscripts related to the Ramayana and Buddhist texts can foster stronger cultural and

diplomatic bonds with countries like Thailand, Cambodia, Indonesia, and Myanmar, where these epics are an integral part of their national identity.

- **Building bridges with the Middle East and Central Asia:** The shared history of trade and knowledge exchange is documented in manuscripts. Highlighting these connections can strengthen contemporary diplomatic relations.
- **Engagement with the Global South:** By collaborating with nations in Africa and South America on manuscript conservation and digitization, India can share its expertise and foster a sense of shared heritage. This positions India as a leader in cultural preservation, not just for its own history but for the global community.



Working Group VIII

**Theme: Legal and Ethical Frameworks for Manuscript Preservation
and Access**

Coordinator

Prof. Richa Kamboj
HoD, Kaladarsana Division, IGNCA

Co-Coordinator

Shri Shah Faesal
Deputy Secretary, Ministry of Culture, Govt. of India

Experts

Shri Safir Anand
Senior Partner & Head of Department (Trademarks, Contractual &
Commercial IP), Anand & Anand Law Firm
Prof. Unnat P. Pandit
Controller General of Patents, Design & Trade Marks, RoC, GI & SICLDR
Prof. Alka Chawla
Professor, Campus Law Centre, Faculty of Law, University of Delhi
Prof. Prabuddha Ganguli
CEO, Vision-IPR
Shri Jagdish Swaroop
Deputy Registrar of Copyright (Retd) and Legal Consultant, M/o Chemical
and Fertilisers
Dr. G. R. Raghavender
Senior Consultant (IPR) & Former Joint Secretary, Govt. of India

Reclaiming India's Knowledge Legacy Through Manuscript Heritage

Introduction

The world's greatest collection of rare manuscripts may be found in India. These manuscripts are a valuable source of the ancient knowledge system of the country, encompassing literature, mathematics, music, religion, science, and law, among others. These manuscripts, which were written in a variety of languages and scripts, offer glimpses of the magnificent Indian cultural history that dates back several centuries. In this digital era, libraries have incorporated technologies into their daily services. The libraries are doing commendable work by digitizing these manuscripts. However, manuscripts remain a primary source of information. The manuscripts must be preserved for future utilization, and their safety and accessibility must be ensured. As a matter of fact, the manuscripts are a great national asset.

Need of Digitization of Manuscripts

1. In the area of data storage, digital technology has revealed a fresh viewpoint. There are millions of websites on the World Wide Web, and information gathering, sharing, expression, publication, and research are now completely integrated into and totally reliant upon the internet. A large number of libraries at present have shifted to developing digital substitutes from their existing resources in addition to maintaining and offering access to 'born digital material'. Digital information that is standardized, organized, and readily available on demand is acquired, converted, stored, and provided through digitization. To prevent the deterioration of manuscripts, it is important to store them in digital files. Manuscripts must be digital and preserved in this digital era. Digitization provides a way to preserve their context without risking damage to the original copies. For digital files of manuscripts, it is easy to search the database of context and quickly and easily find specific keywords to save time. Since there is no risk of tearing or fading the original manuscripts, digitization makes them more accessible to users.

2. The digitization and preservation of manuscripts need some key components: -

- Rescue: First, rescue the original manuscripts to save the documents' lives;
- Scanning: Scanning the original physical manuscripts by high-resolution scanners;
- Image Processing: After scanning, the image is processed to enhance its quality.

- **Metadata Creation:** Metadata is descriptive information about a document. So, it is searchable by any type of keywords.
- **Optical Character Recognition (OCR):** OCR is a software that is used to convert scanned text and images into searchable and editable text.
- **Digital Preservation:** It involves storing the digital files of manuscripts, and
- **Online Access:** It is made available online through a website or digital library so that one just can easily access it at anytime, anywhere.

3. However, still, the archiving of manuscripts can face many challenges and issues, like:

- **Lack of standardization:** since manuscripts are old documents, there is a problem with maintaining the standardization of the digitization of manuscripts. It is difficult to understand spelling, punctuation, abbreviations, etc..
- **Multiple languages:** Manuscripts written in many languages and scripts add further complexity to the archiving process.
- **Large Volume Data:** Sometimes the manuscripts are so large in size that they make it difficult to archive.
- **Copyright Ownership Issues:** Digitization of original documents in manuscripts raises issues of copyright and ownership.

For security reasons, embedding digital watermarks into digital copies can help identify and track unauthorized use.

Legal Framework for Digitization

4. The protection of manuscripts in India is addressed through a combination of legal provisions and dedicated initiatives like the National Mission of Manuscript (NMN), now known as Gyan Bharatam Mission. The legal framework includes the Constitution of India, which emphasizes the preservation of cultural heritage, as well as specific Acts such as the Copyright Act, 1957, which protects the copyrights of the creator of a particular work, and the Ancient Monuments and Archaeological Sites and Remains Act, 1958, which, among other things, protects manuscripts.

5. Article 29 of the Constitution of India protects the rights of citizens to conserve their distinct languages, scripts, or culture. Article 51A (F) makes it a fundamental duty to value and preserve India's rich composite culture.

Copyright/Intellectual Property Rights (IPR) are major legal factors in digital preservation. Digitization involves copying, which may require rights clearance, licenses, or statutory exceptions/fair uses to avoid infringement. While archives often have rights to digitize materials in their possession, challenges still exist with licensing and preserving 'born-digital' content. Digital Rights Management (DRM) can also pose preservation challenges. 'Fair Use' exceptions may apply in certain contexts, and licensing schemes address issues with untraceable rights holders. Understanding when a work enters the public domain is crucial, as it can be freely conserved and accessed. However, it is worth noting that even after copyrights expire, authors may retain moral rights, such as the right to claim authorship and to prevent the destruction or mutilation of their work. Apart from provisions contained in national copyright law, the copyright laws vary globally, which is a factor in international digitization projects. 'Orphan works,' however, pose a challenge in the digitization of works. To overcome it, the creation of a central database for works where the rights holder is unknown or untraceable is very helpful. This would allow institutions /individuals to seek licenses after conducting diligent searches, thereby reducing the risks associated with using such works. Another step may be required to implement a statutory licensing mechanism for orphan works, specifically for non-commercial uses, with fair compensation provisions if the right holder eventually comes forward.

6. Cultural heritage laws and conventions are other important factors which are must to be followed in digitization projects like UNESCO convention which aims at protecting and preserving cultural heritage globally promoting cooperation and encouraging the return of cultural heritage to its country of origin, UNIDROIT convention which focuses on the restitution of stolen or illegally exported cultural objects-establishing rules for their return and promoting ethical standards in cultural heritage management and national laws of respective countries who have their own law governing the ownership, export and preservation of cultural heritage including manuscripts- like India's own Copyright Act, 1957 which defines the scope of copyright and exceptions.

Ethical Framework

7. The digitization of culturally sensitive archaeological materials introduces several ethical concerns that must be addressed to ensure respect and fairness for the resource communities/individuals, which may include: -

(i)cultural appropriation and misrepresentation- digitized works can be misused or misrepresented when taken out of their cultural context. Without proper guidelines, digital productions may lead to cultural misappropriation, where elements of a community's heritage are used in ways that may be disrespectful or exploitative.

(ii)ownership and control- the issue of who owns and controls digitized cultural materials is complex. Traditional custodians often have specific protocols and restrictions regarding their cultural heritage. Digitization efforts must recognize and respect these rights, giving resource communities a role in the decision-making process and control over how their heritage is shared and used.

(iii) Consent and participation- obtaining informed consent from resource communities individuals is a fundamental ethical consideration. Digitization projects should involve these resource communities/individuals from the outset, ensuring that they have a say in how their cultural materials are handled, represented, and distributed.

(iv) Cultural Sensitivity and Privacy: Some cultural/religious materials may be considered sacred or private and may not be suitable for public dissemination. Ethical digitization practices involve respectfully respecting these limitations and ensuring that sensitive materials are treated with the utmost confidentiality. Researchers and institutions must be alert to and adhere to the cultural protocols, particularly those within, and avoid making them publicly accessible if it contravenes the wishes of the resource community.

(v) Impact on Community Identity: The digitization and widespread dissemination of cultural materials can impact the identity and traditions of source communities. There is a risk that the digital presence of these materials may overshadow or alter traditional practices. Ethically, digitization projects should consider the potential effects on community cohesion and cultural integrity.

8. Fostering ethical dialogue: Adding the above-stated ethical issues requires ongoing dialogue and collaboration between researchers, institutions, and resource communities to ensure that digitization efforts are conducted in a manner that is respectful, equitable, and culturally sensitive.

9. Balancing preservation and access is an important factor, as the legal and ethical framework for Manuscript conservation and access is designed to strike a balance between the need to preserve these valuable resources for future generations and the desire to make them accessible for research, education, and cultural enrichment. This delicate balance requires careful consideration of copyright, cultural heritage, and other applicable laws, data protection, and ethical guidelines. Furthermore, it is essential to adopt best

practices that foster respect, collaboration, and compliance with applicable laws.

Review of Relevant Laws in Existence in India and Further Modification/ Amendments Required:

10. As a matter of fact, India has a robust legal and programmatic framework. Its distinct manuscript heritage, however, is not well exploited. There is considerable scope to make our legal framework more foolproof, effective, and well-equipped to handle the emergent challenges of the newly emerging challenges in this digital era. Whereas the existing copyright framework, under the Copyright Act, 1957, in India, should serve as the fundamental pillar for protecting original manuscripts, the other relevant laws may also need a relook and a harmonious approach. Securing access to Indian manuscripts held in foreign museums requires a multifaceted legal framework that depends on national law, international agreements, and collaborative institutions. Some contradictions need to be addressed, such as the Antiquities and Art Treasures Act, 1972, which protects movable cultural properties, including manuscripts, if they are at least 100 years old. The proposed National Manuscript Bill, 2023, limits the scope to seventy-five years in Section 2 of the Bill.

11. Most of the current global Copyright Laws, including Indian copyright law, face challenges in balancing the need to protect original manuscripts and ensuring public access, particularly in the digital age. Some modifications and amendments may be suggested on the following lines:

- Expanded Scope for non-profit and educational uses: Explicitly allow libraries, archives, and educational institutions greater flexibility in digitizing, preserving, and providing limited access to copyrighted manuscripts for educational purposes, research, and other non-commercial purposes akin to provisions for disabled persons.
- Clarification of purposes and character of use in the digital realm: Provide clearer guidelines on how fair use applies to preservation activities like migrating formats or creating access copies, especially regarding born-digital content. This would reduce legal uncertainty for institutions.
- Promoting open access repositories: Encourage and provide incentives for authors/publishers to use legal tools to deposit manuscripts in institutional repositories, ensuring long-term access and preservation in their own intent as well.

- Establish a well-equipped Orphan Work Registry: A central database can be created for copyrighted works where the right holders are unknown or non-traceable. This would allow institutions to seek licences after diligent searches, reducing risks associated with such works.
- Facilitating reasonable licensing schemes: Implement statutory licensing mechanisms for orphan works, especially for non-commercial uses, with fair compensation provisions if the right holder eventually comes forward.
- Exploring blockchain and watermarks for Digital Rights Management (DRM): Investigate and potentially implement technologies that can embed identifying information within digital manuscripts, helping in evidentiary tracking and discouraging unauthorized use.
- Developing standardized metadata for digital preservation: Encourage consistent use of metadata standards that accurately describe and track digital manuscripts' origins, versions, and preservation actions.
- Strengthening international treaties to address new forms of digital authorship: update and adapt international treaties like Berne Convention, WCT, WPPT, etc., to better address new forms of digital authorship and dissemination that fall outside traditional categories like books and music.
- Educating authors about their rights and options: raising awareness amongst authors about their rights, including the ability to negotiate licensing terms, retain certain rights, and deposit works in repositories in their own interest.
- Providing resources for authors to protect their works: offer accessible resources and guidance for authors/owners to register their copyrights and enforce their rights effectively, particularly in the digital landscape.

These improvements require careful consideration by the Government and work for better collaboration between author/owner, publishers, libraries, archives, technology providers and legal/domain experts to achieve a balanced and effective copyright framework for meaningful protection and access in the digital age.

Prominent Case Laws in India on Manuscript Protection:

Here are some Prominent case laws that, while not exclusively on manuscripts, establish principles relevant to their protection in India:

(i) Najma Heptulla v. Orient Longman Ltd. & Ors. (1988)

This is a very significant case directly related to a manuscript. The dispute revolved around the unpublished portion of Maulana Abul Kalam Azad's autobiography, "India wins freedom." The case dealt with the ownership of copyright and the right of the legal heirs to control its publication. The Delhi High Court's decision affirmed that the copyright in the manuscript belonged to the author's legal heir, and the publisher's rights were subject to the agreement made with the author. This case highlights the principle that the transfer of a physical manuscript does not automatically transfer the copyright to the manuscript.

(ii) Eastern Book Co. & others vs D.B. Modak & Another (2008)

The landmark Supreme Court case is crucial for understanding the concept of "originality" under Indian copyright law. The case involved the copyright of headnotes and other editorial inputs in published judgments. The court held that even though the judgments themselves are in the public domain, the editorial work that involves a "modicum of creativity" is protected by copyright. This principle can be applied to manuscripts where any original and notations, compilations, or scholarly additions made to a public domain text could be eligible for copyright protection.

Acquisition of Manuscripts as Heritage Property of National Importance

The core issue is whether and how the Government of India can legally acquire manuscripts of national importance from private entities or institutions for preservation, while balancing such acquisition with constitutional guarantees of property rights and cultural obligations. The National Manuscripts Bill, 2023, envisions the establishment of a National Manuscripts Authority (NMA) with the authority to acquire, preserve, and regulate manuscripts of historical, cultural, and artistic significance. The proposed NMA, as outlined in the National Manuscripts Bill, 2023, is slated to have proposed powers, enabling it to purchase or compulsorily acquire manuscripts from private collectors based on their uniqueness or vulnerability. The legal authority of the NMA to purchase or compulsorily acquire manuscripts from private custody must be firmly grounded in constitutional and statutory principles.

❖ Constitutional Provisions

The Constitution of India incorporates the provisions of Articles 49 and 51(f) to protect cultural heritage. The right to conserve, protect, and manage cultural heritage is not included in the Fundamental Rights of the Indian Constitution; instead, it is accorded a non-enforceable status.

Article 49 stipulates that it shall be the obligation of the State to protect every monument or place or object of artistic or historic interest, declared by or under law made by Parliament to be of national importance, from spoliation, disfigurement, destruction, removal, disposal or export, as the case may be. **Article 51A(f)** stipulates that it shall be the duty of every citizen of India to value and preserve the rich heritage of our composite culture.

Thus, “object of artistic or historic interest” comfortably covers manuscripts of “historical, cultural, artistic or scientific value,” and Parliament has in fact enacted laws that **expressly include manuscripts** meeting age/value thresholds within “antiquities.”

Apart from the above, **Entry 67 of List I of the Seventh Schedule** endorses the jurisdiction of the ancient and historical monuments and records, and archaeological sites and remains, declared by or under law made by Parliament to be of national importance, and **Entry 12 of List II** incorporates the protection of libraries, museums and other similar institutions controlled or financed by the State; ancient and historical monuments and records other than those declared by or under law made by Parliament to be of national importance.

❖ Case Law : Scope of Protection of Intellectual Property under Article 300-A: *K.T. Plantation (P) Ltd. v. State of Karnataka* [(2011) 9 SCC 1]

Article 300-A proclaims that no person can be deprived of his property save by authority of law, meaning thereby that a person cannot be deprived of his property merely by an executive fiat, without any specific legal authority or without the support of law made by a competent legislature. The expression “property” in Article 300-A is not confined to land alone; it includes intangibles like copyrights and other intellectual property and embraces every possible

interest recognised by law. Article 300-A, therefore, protects private property against executive action.

The principles of eminent domain, as such, are not seen incorporated in Article 300-A, as we see in Article 30(1-A), as well as in the second proviso to Article 31-A(1), though we can infer those principles in Article 300-A. The provision for payment of compensation has been specifically incorporated in Article 30(1-A) as well as in the second proviso to Article 31-A(1) for achieving specific objectives. The Constitution (Forty-fourth Amendment) Act, 1978, while omitting Article 31, brought in a substantive provision clause (1-A) to Article 30. Resultantly, though no individual or even educational institution belonging to majority community shall have any fundamental right to compensation in case of compulsory acquisition of his property by the State, an educational institution belonging to a minority community shall have such fundamental right to claim compensation in case the State enacts a law providing for compulsory acquisition of any property of an educational institution established and administered by a minority community. Further, the second proviso to Article 31-A(1) prohibits the legislature from making a law which does not contain a provision for payment of compensation at a rate not less than the market value, so that a law which does not contain such a provision shall be invalid, and the acquisition proceedings would be rendered void.

❖ **Public Purpose**

Deprivation of property within the meaning of Article 300-A, generally speaking, must take place for a public purpose or public interest. The concept of eminent domain, which applies when a person is deprived of his property, postulates that the purpose must be primarily public and not primarily of private interest and merely incidentally beneficial to the public. Any law that deprives a person of his private property for private interest will be unlawful and unfair, undermining the rule of law, and can be subjected to judicial review. But the question as to whether the purpose is primarily public or private has to be decided by the legislature, which, of course, should be made known.

The concept of public purpose has been given a fairly expansive meaning, which has to be justified upon the purpose and object of the statute and the policy of the legislation. Public purpose is, therefore, a condition precedent for invoking Article 300-A.

❖ Compensation

The requirement of public purpose is invariably the rule for depriving a person of his property, violation of which is amenable to judicial review. Let us now examine whether the requirement of payment of compensation is the rule after the deletion of Article 31(2). Payment of compensation amount is a constitutional requirement under Article 30(1-A) and under the second proviso to Article 31-A(1), unlike Article 300-A. After the Forty-fourth Amendment Act, 1978, the constitutional obligation to pay compensation to a person who is deprived of his property primarily depends upon the terms of the statute and the legislative policy. Article 300-A, however, does not prohibit the payment of just compensation when a person is deprived of their property. The question is whether a person is entitled to receive compensation, as a matter of right, in the absence of any stipulation in the statute that deprives them of their property.

In addition to the Constitutional provisions, existing statutes already supply acquisition and access levers: the **Antiquities and Art Treasures Act, 1972**, notably **Section 19 (compulsory acquisition)** with **Section 20 (compensation/arbitration)**, and the **Public Records Act, 1993**, in particular **Sections 3, 11, 14, 17 (acceptance from private sources; directions to acquire; purchase/lease; power to make rules)**. These provide immediate, defensible pathways to purchase, require registration, and, in limited cases, compulsorily acquire manuscripts.

❖ Other Statutory Provisions

1. Antiquities and Art Treasures Act, 1972

This AATA regulates movable heritage, including museum collections, artifacts, and manuscripts. Preserving heritage is essential, as it enables future generations to understand and appreciate their cultural roots, thereby helping to build a sense of pride and identity. This Act was enacted with the object of regulating the export, trade, and preservation of antiquities and art treasures. Section 2(1)(a) defines “antiquity” to include manuscripts of historical interest which are not less than one hundred years old. The Act empowers the Central Government to compulsorily register antiquities and regulate their transfer. However, the Act does not contain any direct provision for the acquisition of

antiquities from private ownership, except in cases of unlawful possession, smuggling, or violation of licensing conditions.

Section 2. Definitions.

(a) “antiquity” includes—

(1) (i) any coin, sculpture, painting, epigraph or other work of art or craftsmanship; (ii) any article, object or thing detached from a building or cave; (iii) any article, object or thing illustrative of science, art, crafts, literature, religion, customs, morals or politics in bygone ages; (iv) any article, object or thing of historical interest; (v) any article, object or thing declared by the Central Government, by notification in the Official Gazette, to be an antiquity for the purposes of this Act, which has been in existence for not less than one hundred years; and **any manuscript, record or other document which is of scientific, historical, literary or aesthetic value and which has been in existence for not less than seventy-five years;**

Thus, manuscripts fall squarely within the definition of “antiquity.”

Section 14. Registration of antiquities.

(1) The Central Government may, from time to time, by notification in the Official Gazette, specify those antiquities which shall be registered under this Act.

(2) In specifying the antiquities under sub-section (1), the Central Government shall have regard to the following factors, namely:— (i) the necessity for conserving the objects of art; (ii) the need to preserve such objects within India for the better appreciation of the cultural heritage of India; (iii) such other factors as will, or are likely to, contribute to the safeguarding of the cultural heritage of India.

(3) Every person who owns, controls or is in possession of any antiquity specified in the notification issued under sub-section (1) shall register such antiquity before the registering officer— (a) in the case of a person who owns, controls or possesses such antiquity on the date of issue of such notification, within three months of such date; and (b) in the case of any other person, within fifteen days of the date on which he comes into ownership, control or possession of such antiquity, and obtain a certificate in token of such registration.

Section 16. Application for Registration and Grant of Certificate of Registration.

(1) Every person required to register any antiquity before the registering officer under section 14 shall make an application to the registering officer for the grant of a certificate of registration.

(2) Every application under sub-section (1) 1 [shall, in the case of such antiquities or class of antiquities as the Central Government may, by notification in the Official Gazette, specify, be accompanied] by such photographs of the antiquity which is to be registered and by such number of copies, not exceeding six, as may be prescribed and shall be made in such form and shall contain such particulars as may be prescribed.

(3) On receipt of an application under sub-section (1), the registering officer may, after holding such inquiry as he deems fit, grant a certificate of registration containing such particulars as may be prescribed.

(4) No application made under this section shall be rejected unless the applicant has been given a reasonable opportunity of being heard in the matter.

Section 17. Transfer of ownership, etc., of antiquities to be intimated to the registering officer.

Whenever any person transfers the ownership, control or possession of any antiquity specified in any notification issued under sub-section (1) of section 14 such person shall intimate, within such period and in such form as may be prescribed, the fact of such transfer to the registering officer.

Section 19. Power of the Central Government to compulsorily acquire antiquities and art treasures.

(1) If the Central Government is of the opinion that it is desirable to preserve any antiquity or art treasure in a public place, that Government may make an order for the compulsory acquisition of such antiquity or art treasure.

(2) On the making of an order under sub-section (1) the Collector of the district in which such antiquity or art treasure is kept shall give notice to the owner thereof intimating him of the decision of the Central Government to acquire the same and it shall be lawful for the Collector to take possession of such antiquity or art treasure, for which purpose the Collector may use such force as may be necessary.

(3) Where the owner of any antiquity or art treasure the possession of which has been taken over by the Collector under sub-section (2) objects to the

taking over of such possession, he may, within a period of thirty days from the date on which such possession was taken over, make a representation to the Central Government putting forth his objections: Provided that the Central Government may entertain the representation after the expiry of the said period of thirty days, if it is satisfied that the owner of such antiquity or art treasure was prevented by sufficient cause from making the representation in time.

(4) On receipt of any representation under sub-section (3), the Central Government, after making such inquiry as it deems fit and after giving to the objector an opportunity of being heard in the matter, shall, within a period of ninety days from the date of receipt of the representation, either rescind or confirm the order made by it under sub-section (1).

(5) Where any order made by the Central Government under sub-section (1) is rescinded under sub-section (4) the antiquity or art treasure shall be returned to the owner thereof without delay and at the expense of the Central Government.

(6) Where the order made by the Central Government under sub-section (1) is confirmed under sub-section (4) the antiquity or art treasure shall vest in the Central Government with effect from the date on which the possession thereof has been taken over by the Collector under sub-section (2).

(7) The power of compulsory acquisition conferred by this section shall not extend to any object, being an antiquity or art treasure, used for bona fide religious observances.

Explanation. —In this section, “public place” means any place which is open to the use of the public, whether on payment of fees or not, or whether it is actually used by the public or not.

This enables the compulsory acquisition to preserve any antiquity or art treasure in a public place, which in its fold, can be read to include manuscripts of national importance as the definition of the term “antiquity” as per Section 2(a) of the Act includes “manuscripts”.

Section 20. Payment of compensation for antiquities and art treasures compulsorily acquired under section 19.

(1) Where any antiquity or art treasure is compulsorily acquired under section 19, there shall be paid compensation, the amount of which shall be determined in the manner and in accordance with the principles hereinafter set out, that is to say,— (a) where the amount of compensation can be fixed

by agreement, it shall be paid in accordance with such agreement; (b) where no such agreement can be reached, the Central Government shall appoint as arbitrator a person who is, or has been, or is qualified for appointment as, a Judge of a High Court; (c) the Central Government may, in any particular case, nominate a person having expert knowledge as to the nature of the antiquity or art treasure compulsorily acquired to assist the arbitrator and where such nomination is made, the person to be compensated may also nominate an assessor for the same purpose; (d) at the commencement of the proceedings before the arbitrator, the Central Government and the person to be compensated shall state what, in their respective opinion, is a fair amount of compensation; (e) the arbitrator shall, after hearing the dispute, make an award determining the amount of compensation which appears to him to be just and specifying the person or persons to whom such compensation shall be paid and in making the award he shall have regard to the circumstances of each case and the provisions of sub-section (2); (f) where there is any dispute as to the person or persons who are entitled to the compensation, the arbitrator shall decide such dispute and if the arbitrator finds that more persons than one are entitled to compensation, he shall apportion the amount thereof amongst such persons; (g) nothing in the Arbitration Act, 1940 (10 of 1940) shall apply to arbitration under this section.

(2) While determining the compensation under sub-section (1), the arbitrator shall have regard to the following factors, namely:— (i) the date or the period to which the antiquity or art treasure belongs; (ii) the artistic, aesthetic, historical, architectural, archaeological or anthropological importance of the antiquity or art treasure; (iii) the rarity of the antiquity or art treasure; (iv) such other matters as are relevant to the dispute.

(3) The arbitrator appointed under sub-section (1), while holding arbitration proceedings under this section, shall have all the powers of a Civil Court, while trying a suit, under the Code of Civil Procedure, 1908 (5 of 1908), in respect of the following matters, namely:— (a) summoning and enforcing the attendance of any person and examining him on oath; (b) requiring the discovery and production of any document; (c) reception of evidence on affidavits; (d) requisitioning any public record from any court or office; (e) issuing commissions for the examination of witnesses.

Section 23. Powers of entry, search, seizure, etc.

(1) Any person, being an officer of Government, authorized in this behalf by the Central Government, may, with a view to securing compliance with the provisions of this Act or to satisfying himself that the provisions of this Act have been complied with—

(i) enter and search any place;

(ii) seize any antiquity or art treasure in respect of which he suspects that any provision of this Act has been, is being, or is about to be, contravened and thereafter take all measures necessary for securing the production of the antiquity or art treasure so seized in a court and for its safe custody, pending such production.

(2) The provisions of sections 102 and 103 of the Code of Criminal Procedure, 1898 (5 of 1898) relating to search and seizure shall, so far as may be, apply to searches and seizures under this section.

Section 24. Power to determine whether or not an article, etc., is an antiquity or an art treasure.

If any question arises whether any article, object or thing or manuscript, record or other document is or is not an antiquity or is or is not an art treasure for the purposes of this Act, it shall be referred to the Director General, Archaeological Survey of India, or to an officer not below the rank of a Director in the Archaeological Survey of India authorized by the Director General, Archaeological Survey of India and the decision of the Director General, Archaeological Survey of India or such officer, as the case may be, on such question shall be final.

2. Ancient Monuments and Archaeological Sites and Remains Act, 1958

The AMASR Act primarily concerns itself with monuments and archaeological sites. Section 3(d) defines “antiquity” in terms comparable to the AATA, but the acquisition powers under the Act (notably Section 19 for compulsory acquisition of antiquities) are restricted to antiquities discovered in the course of excavation or found at protected sites. The scope of acquisition does not presently extend to manuscripts held by private custodians.

To extend the AMASR framework to manuscripts, amendments would be necessary to broaden the definition of antiquity expressly to cover manuscripts of cultural or historical importance, irrespective of age, and to introduce an enabling provision empowering the Government to compulsorily acquire such manuscripts on payment of just compensation. Such an amendment would harmonize with Article 49 and could be defended on the grounds of preserving national heritage.

Section 2(b) defines “antiquity” to include— (i) any coin, sculpture, manuscript, epigraph, or other work of art or craftsmanship, (ii) any article, object or thing detached from a building or cave, (iii) any article, object or thing illustrative of science, art, crafts, literature, religion, customs, morals or politics in bygone ages, (iv) any article, object or thing of historical interest, and (v) any article, object or thing declared by the Central Government, by notification in the Official Gazette, to be an antiquity for the purposes of this Act, which has been in existence for not less than one hundred years.

The Ancient Monuments and Archaeological Sites and Remains Act, 1958 (“the AMASR Act”), was enacted with the primary objective of safeguarding immovable heritage, namely ancient monuments, archaeological sites, and remains of historical significance. While the definitional clause under Section 2(b) does include “manuscripts” within the broader category of “antiquity”, specifically, clause (i) encompasses “any coin, sculpture, manuscript, epigraph, or other work of art or craftsmanship... which has been in existence for not less than one hundred years”: the operative provisions of the statute remain structurally focused on immovable cultural property. The acquisition provisions, as contained in Sections 5, 6, 13, and 14, are explicitly tailored towards “protected monuments” or “archaeological sites” and not towards portable cultural objects such as manuscripts. Accordingly, while manuscripts are covered within the definitional scope of “antiquities,” the Act does not presently contain a self-sufficient framework to treat manuscripts as a distinct class of cultural heritage subject to compulsory acquisition, guardianship, or custodial regulation.

○ **Whether Manuscripts Can be Acquired Under Current Law**

On a plain reading of the AMASR Act, it becomes evident that the provisions relating to compulsory acquisition are confined to “protected monuments” and do not extend to movable antiquities such as manuscripts. Sections 13 and 14 empower the Central Government to acquire a protected monument under specified circumstances, but no analogous mechanism is provided for antiquities of a portable nature. Consequently, manuscripts of national importance cannot be compulsorily acquired under the present statutory scheme of the AMASR Act. For such purposes, reliance is instead placed upon other legislation, particularly the Antiquities and Art Treasures Act, 1972, which specifically regulates the movement, ownership, and export of antiquities, including manuscripts. The inclusion of manuscripts within

the definitional ambit of “antiquities” under Section 2(b) of the AMASR Act, therefore, serves a classificatory function without conferring operative acquisition powers in respect thereof.

○ **Proposed Amendments to Enable Acquisition of Manuscripts**

If the legislative intent is to bring manuscripts fully within the protective and acquisition framework of the AMASR Act, targeted amendments would be required.

(A) Definitions. A distinct statutory definition of “manuscript” ought to be inserted, such as: “*‘Manuscript’ means any handwritten or printed document, record, text, or compilation of historical, cultural, religious, or artistic value, irrespective of the medium (paper, palm leaf, birch bark, parchment, cloth, or other material), existing for not less than seventy-five years, or as may be notified by the Central Government.*” Furthermore, manuscripts should be expressly recognized as a sub-category of antiquities, subject to a protective regime.

(B) Declaration of National Importance. The scheme of Sections 3 and 4, which presently applies to monuments and sites, should be extended mutatis mutandis to cover manuscripts and other documentary heritage. A notification procedure for declaring manuscripts to be of “national importance” should be incorporated.

(C) Custody and Acquisition. A new chapter entitled “*Protection and Preservation of Manuscripts of National Importance*” should be introduced, enabling: (i) voluntary transfer of manuscripts to the Government or a designated National Manuscripts Authority by way of gift or sale; (ii) compulsory acquisition of manuscripts by the Central Government, analogous to Section 13, with preservation expressly recognised as a “public purpose” as under the Land Acquisition framework; and (iii) a guardianship model, under which ownership may remain with the private custodian, but the manuscript is placed under statutory custodial supervision in designated repositories.

(D) Institutional Mechanism. The statute should establish or empower a National Manuscripts Authority, with responsibility for acquisition, preservation, and custodianship of manuscripts declared to be of national importance. Provision should also be made for the creation and maintenance of a national repository of manuscripts with statutory backing.

3. Public Records Act, 1993

The PRA primarily governs public records maintained by government offices, public sector undertakings, and statutory bodies. Section 2(e) defines “public records” to mean records created by such authorities. Manuscripts in the possession of private individuals, trusts, or religious bodies fall outside its scope. The Act authorises the Director General of Archives to regulate the management, preservation, and destruction of public records, but does not authorise acquisition from private sources.

Therefore, to utilise this framework, the PRA would require a substantial amendment expanding the definition of “public records” to include private manuscripts of national importance, coupled with an express provision empowering the Government to acquire such records in the public interest. Constitutional support may be drawn from Entry 12 of List I (Union control over archives, ancient and historical records) read with Article 49.

The relevant provisions of PRA include:

Section 2 (e) “public records” includes

(i) any document, manuscript, and file;

(ii) any microfilm, microfiche, and facsimile copy of a document;

(iii) any reproduction of image or images embodied in such microfilm (whether enlarged or not); and

(iv) any other material produced by a computer or by any other device, of any records-creating agency;

Section 3. Power of the Central Government to coordinate, regulate, and supervise operations connected with the administration, management, etc., of public records.

(1) The Central Government shall have the power to coordinate, regulate, and supervise the operations connected with the administration, management, preservation, selection, disposal, and retirement of public records under this Act.

(2) The Central Government in relation to the public records of the records creating agencies specified in sub-clauses (i) and (ii) of clause (f) of section 2 and the Union territory Administration in relation to the public records of

the records creating agencies specified in sub-clauses (iii) and (iv) of the said clause, may, by order, authorise the Director General or the head of the Archives, as the case may be, subject to such conditions as may be specified in the order, to carry out all or any of the following functions, namely:—

(a) supervision, management, and control of the Archives.

(b) acceptance for deposit of public records of a permanent nature after such period as may be prescribed;

(c) custody, use, and withdrawal of public records;

(d) arrangement, preservation, and exhibition of public records;

(e) preparation of inventories, indices, catalogues, and other reference media of public records;

(f) analysing, developing, promoting, and coordinating the standards, procedures, and techniques for improvement of the records management system;

(g) ensuring the maintenance, arrangement, and security of public records in the Archives and in the offices of the records-creating agency;

(h) promoting utilisation of available space and maintenance of equipment for preserving public records;

(i) tendering advice to records creating agencies on the compilation, classification, and disposal of records and application of standards, procedures, and techniques of records management;

(j) survey and inspection of public records;

(k) organising training programmes in various disciplines of Archives administration and records management;

(l) accepting records from any private source;

(m) regulating access to public records;

(n) receiving records from defunct bodies and making arrangements for securing public records in the event of a national emergency;

(o) receiving reports on records management and disposal practices from the records officer;

(p) providing authenticated copies of, or extracts from, public records;

(q) destroying or disposal of public records;

(r) obtaining on lease, purchasing, or accepting as a gift any document of historical or national importance.

Section 11. Receipt of records from private sources.

(1) The National Archives of India or the Archives of the Union territory may accept any record of historical or national importance from any private source by way of gift, purchase or **otherwise**.

Section 14. Functions of the Board.

The Board shall perform the following functions, namely:

(a) Advise the Central Government and Union territory Administrations on matters concerning the administration, management, conservation, and use of public records;

(b) Lay down guidelines for training of Archivists;

(c) Give directions for the acquisition of records from private custody.

(d) Deal with such other matters as may be prescribed.

Section 17. Power to make rules.

(1) The Central Government may, by notification in the Official Gazette, make rules to carry out the provisions of this Act.

(2) In particular and without prejudice to the generality of the foregoing power, such rules may provide for all or any of the following matters, namely: —

(a) the period after which public records of permanent nature may be accepted under clause (b) of sub-section (2) of section 3;

(b) the manner in which and the conditions subject to which public records can be destroyed under clause (d) of sub-section (1) of section 6;

(c) The manner in which periodical review of classified public records for downgrading shall be undertaken under clause (f) of sub-section (1) of section 6;

(d) the manner in which the records officer will report to the Director General or the head of the Archives under clause (k) of sub-section (1) of section 6;

(e) the manner in which and the conditions subject to which public records may be destroyed or disposed of under sub-section (1) of section 8;

(f) the manner in which and the conditions subject to which records of historical or national importance may be made available to a research scholar under sub-section (2) of section 11;

- (g) exceptions and restrictions subject to which public records may be made available to a research scholar under sub-section (1) of section 12;
- (h) the manner in which and the conditions subject to which any records creating agency may grant to any person access to public records in its custody under sub-section (2) of section 12;
- (i) the allowances payable to members of the Board under sub-section (3) of section 13;
- (j) the matters with respect to which the Board may perform its functions under clause (d) of section 14;
- (k) any other matter which is required to be, or may be, prescribed.

❖ **The Way Forward**

In their present form, neither the AATA, the PRA, nor the AMASR provides an adequate or direct legal basis for the compulsory acquisition of manuscripts of national importance from private hands. Each would require targeted amendments: (i) the AATA to include explicit acquisition powers over manuscripts declared national treasures; (ii) the PRA to expand its scope to private records of national importance; and (iii) the AMASR Act to widen its acquisition powers beyond monuments to cover manuscripts. Alternatively, the more prudent legislative course would be the enactment of a sui generis Manuscripts Preservation Act, anchored in constitutional principles, and harmonised with the above statutory frameworks.

The Supreme Court in *K.T. Plantation (P) Ltd. v. State of Karnataka* [(2011) 9 SCC 1] observed that deprivation of property under Article 300A must meet the twin tests of (i) public purpose and (ii) compensation, albeit the latter being a legislative policy choice. Acquisition of manuscripts for preservation as national heritage meets the public purpose requirement. The legislative policy should provide for “*just compensation*”, ensuring fairness and possibly insulating the statute from constitutional challenge.

In the case of manuscripts of national importance, their preservation and accessibility for study, research, and dissemination of knowledge directly serve the public purpose requirement. Manuscripts, by their very nature, are not only cultural artefacts but also repositories of civilisational memory, and their conservation facilitates the advancement of education, research, and the

arts. These objectives are explicitly aligned with the Directive Principles of State Policy, particularly Article 49 (protection of monuments and objects of national importance) and Article 51A(f) (fundamental duty of every citizen to value and preserve the rich heritage of our composite culture).

Thus, if the State seeks to acquire manuscripts from institutions, private custodians, or trusts, the justification of providing access to scholars and researchers would squarely satisfy the “public purpose” threshold. The ultimate beneficiary is intended to be the public at large, through the expansion of knowledge and the preservation of culture.

We need to look at the definition of Section 2 of the ANTIQUITIES AND ART TREASURES ACT, 1972, ACT NO. 52 OF 1972 defines an antique as any manuscript, record, or other document which is of scientific, historical, literary, or aesthetic value and which has been in existence for not less than seventy-five years. Starting from this definition, we can create a framework to classify and protect them as antiques. They can then be brought under the law to protect. Access to art can be found in some provisions of the Indian Constitution. Article 29’ of the Constitution: “Any section of the citizens residing in the territory of India or any part thereof having a distinct language, script or culture of its own shall have the right to conserve the same“ ‘Article 49’ of the Indian Constitution: “It shall be the obligation of the State to protect every monument or place or object of artistic or historic interest, (declared by or under law made by Parliament) to be of national importance, from spoilation, disfigurement, destruction, removal, disposal or export, as the case may be. ” ‘Article 51 A(F)’ of the Constitution: “It shall be the duty of every citizen of India to value and preserve the rich heritage of our composite culture”. Our group presentation can address some of these aspects and establish a legal framework for detailed study and recommendations following the conference. A serious study group of experts will need to be commissioned to thoroughly examine the matter with a high level of seriousness, reviewing all facets of the existing laws, critically assessing them, and exploring appropriate amendments to ensure access to manuscripts, including their preservation. Alternatively, the commission may recommend a sui-generis system to address the issues.



OVERALL SUPERVISION & COORDINATION BY

Prof. (Dr.) Ramesh C. Gaur

Dean Administration

Indira Gandhi National Centre for the Arts

ASSISTED BY

Dr. Sarwarul Haque

Asst. Professor, IGNCA

PROOFREADING & SUPERVISION

Dr. Parveen Babbar

Deputy Librarian, JNU

COMPILATION BY

Rahul Pachori, Research Fellow (Linguists)

Kala Nidhi Division, IGNCA

DESIGNED BY

Nikhil K Sharma

Graphic Designer, Sutradhara, IGNCA

Shruti Patel

Graphic Designer (Nodal Cell), Sutradhara, IGNCA

Dheeraj Sharma

Library & Information Assistant, IGNCA



संस्कृति मंत्रालय
Ministry of Culture
Government of India