

SAMATA (समता): A 16-Layer Metadata Application Profile for Cultural and Knowledge Systems

Gopal Adak¹  

¹ Nalanda University; Muses' Attic

Abstract

This poster introduces SAMATA (समता), a layered metadata application profile designed to represent knowledge resources and their transmission across textual, archival, print, oral, and material forms. While standards such as Dublin Core, CIDOC-CRM, and PREMIS provide robust infrastructures for describing information objects, they remain limited in representing collections embedded within complex knowledge systems and community-governed contexts.

Drawing on field-based metadata work across manuscript, archival, print, and museum collections, this study identifies three recurring structural gaps: (1) limited representation of non-Gregorian temporal systems; (2) absence of structured modelling for intellectual transmission relationships; and (3) lack of machine-readable governance metadata addressing community consent and ethical use, including AI training conditions. SAMATA addresses these gaps through a sixteen-layer architecture that extends existing standards while maintaining interoperability. The framework is currently under pilot implementation at Chandernagore College under a formal institutional MoU.

1. Introduction

Metadata standards such as Dublin Core and CIDOC-CRM provide essential infrastructures for interoperability across libraries, archives, and museums. However, these frameworks are primarily oriented toward object description and event modelling, and are less effective in representing the contextual, relational, and governance dimensions of knowledge systems embedded in living communities.

This limitation becomes evident across multiple collection types: manuscript traditions where texts exist within networks of commentary and pedagogical transmission; museum collections where objects carry workshop lineages and iconographic programmes; archival collections shaped by community practices and circulation histories; and oral traditions governed by community authority. SAMATA is proposed as a modular metadata application profile that extends these standards to incorporate transmission, context, and governance within a unified and interoperable structure applicable across these domains.

2. Structural Gaps in Metadata Practice

Three structural gaps emerge consistently from field-based metadata practice:

Temporal Representation: Materials dated using Śaka, Baṅgābda, Vikrama, Hijri, or ritual-lunar systems cannot be represented in structured form within existing schemas. Original temporal expressions are often reduced to approximate Gregorian equivalents, resulting in loss of precision and cultural context.

Poster

Available Aug 01, 2026

Correspondence to
Gopal Adak
cemetry.h@gmail.com



Copyright © 2026 Adak. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Transmission Modelling: Texts, archival records, museum objects, and oral traditions exist within networks of intellectual and material continuity—commentary (*ṭīkā*), recension, workshop succession, guru–shishya, and institutional transmission. Existing standards provide limited mechanisms for encoding these relationships in structured, machine-readable form across collection types.

Governance and Ethical Metadata: Metadata frameworks lack structured fields for representing community governance, consent, and ethical conditions of use. As digitised collections are increasingly incorporated into machine learning workflows, the absence of machine-readable AI Training Eligibility introduces significant limitations for responsible reuse.

3. The SAMATA Framework and Interoperability

SAMATA organises a structured set of fields across sixteen layers grouped into four domains: identification (L1–L4), knowledge and transmission (L5–L7), governance and community (L8–L11), and preservation and extension (L12–L16). It functions as a metadata application profile in which each layer aligns with its closest Dublin Core element or CIDOC-CRM class, with additional SAMATA fields implemented as structured extensions.

Three layers represent core structural contributions: L4 (Temporal Integrity) enables multi-calendar encoding with parallel values and certainty indicators; L5 (Transmission and Intellectual Continuity) models relationships such as *mūla–ṭīkā*, recension, workshop succession, and guru–shishya; and L9 (Human Value and Governance) introduces machine-readable governance fields including AI Training Eligibility (Open / Conditional / Restricted / Sacred).

The framework is modular, supporting partial and progressive implementation across diverse institutional contexts. Preservation components align with PREMIS, while community vocabulary (L10) incorporates folksonomy as a parallel descriptive layer alongside controlled vocabularies.

4. Applied Example and Implementation

SAMATA's architecture operates consistently across collection types. In a commentary manuscript, L5 encodes its position within a transmission network, L4 preserves original calendar expressions, and L9 records community-defined AI Training Eligibility. In a Pala-period bronze, L16 represents iconographic programme and artisan lineage, while L5 encodes workshop succession. The same architecture is adapted across collection types without fragmentation.

The framework is grounded in field-based metadata work across multiple large-scale digitisation datasets exceeding 1.5 million items. Pilot implementation is currently underway at Chandernagore College, West Bengal, under a formal institutional MoU (January 2026), serving as the primary site of empirical testing.

References

- [1] S. R. Carroll and et al., "The CARE Principles for Indigenous Data Governance," *Data Science Journal*, vol. 19, no. 1, p. 43, 2020.
- [2] DCMI, "DCMI Metadata Terms," 2020. [Online]. Available: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- [3] M. Doerr, "The CIDOC Conceptual Reference Model: An ontological approach to semantic interoperability of metadata," *AI Magazine*, vol. 24, no. 3, pp. 75–92, 2003.
- [4] M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
- [5] Local Contexts, "Traditional Knowledge Labels." [Online]. Available: <https://localcontexts.org/labels/traditional-knowledge-labels/>
- [6] H. A. Olson, *The Power to Name*. Kluwer Academic Publishers, 2002.
- [7] M. D. Wilkinson and et al., "The FAIR Guiding Principles," *Scientific Data*, vol. 3, p. 160018, 2016.

UNCORRECTED
PROOF