

# Semantic Mapping of Archival Metadata Standards

Toward a Hybrid Knowledge Organization System  
for the National Library and Archives of Iran

Fatemeh Pazooki (f-pazoki@nlai.ir)

Esmail Babaei Dehkordi (babaeinlai@gmail.com)

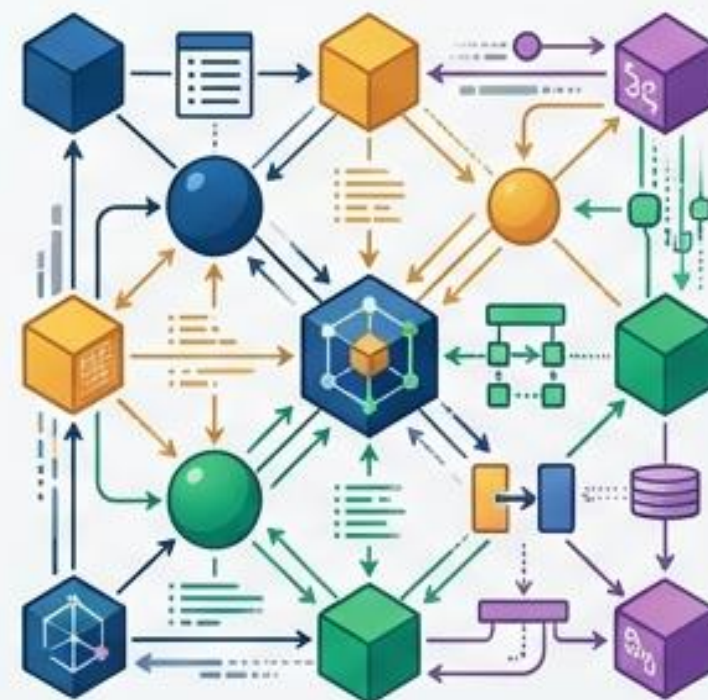
# The Archival Paradigm Shift

## From Physical Repositories



- Static preservation of physical heritage.
- Isolated, institutional-level discovery.

## To Dynamic Networks



- Dynamic, multi-layered hubs for access and semantic reuse.
- Web-based discovery and Linked Data integration.

The Digital Paradox — Unprecedented opportunities for global knowledge sharing are currently blocked by severe semantic complexity when multiple, unintegrated metadata standards coexist.



# The Scale of the Challenge: NLAI Context

By the Numbers

## 4.3+ Million

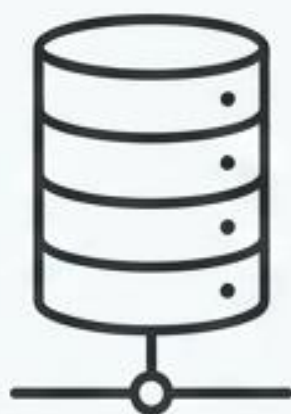
Physical Items: rare manuscripts,  
historical documents, periodicals

## 10.2+ Million

Digitized Pages



1. SANA (Iranian  
Newspapers  
Database)



2. Manuscript  
Database



3. National Archives  
Network

**These distinct systems are governed by heterogeneous metadata standards operating without cohesive semantic alignment.**

# The Fragmentation of Archival Context

## RiC-CM

Focuses on context and relational archival entities.

## PREMIS

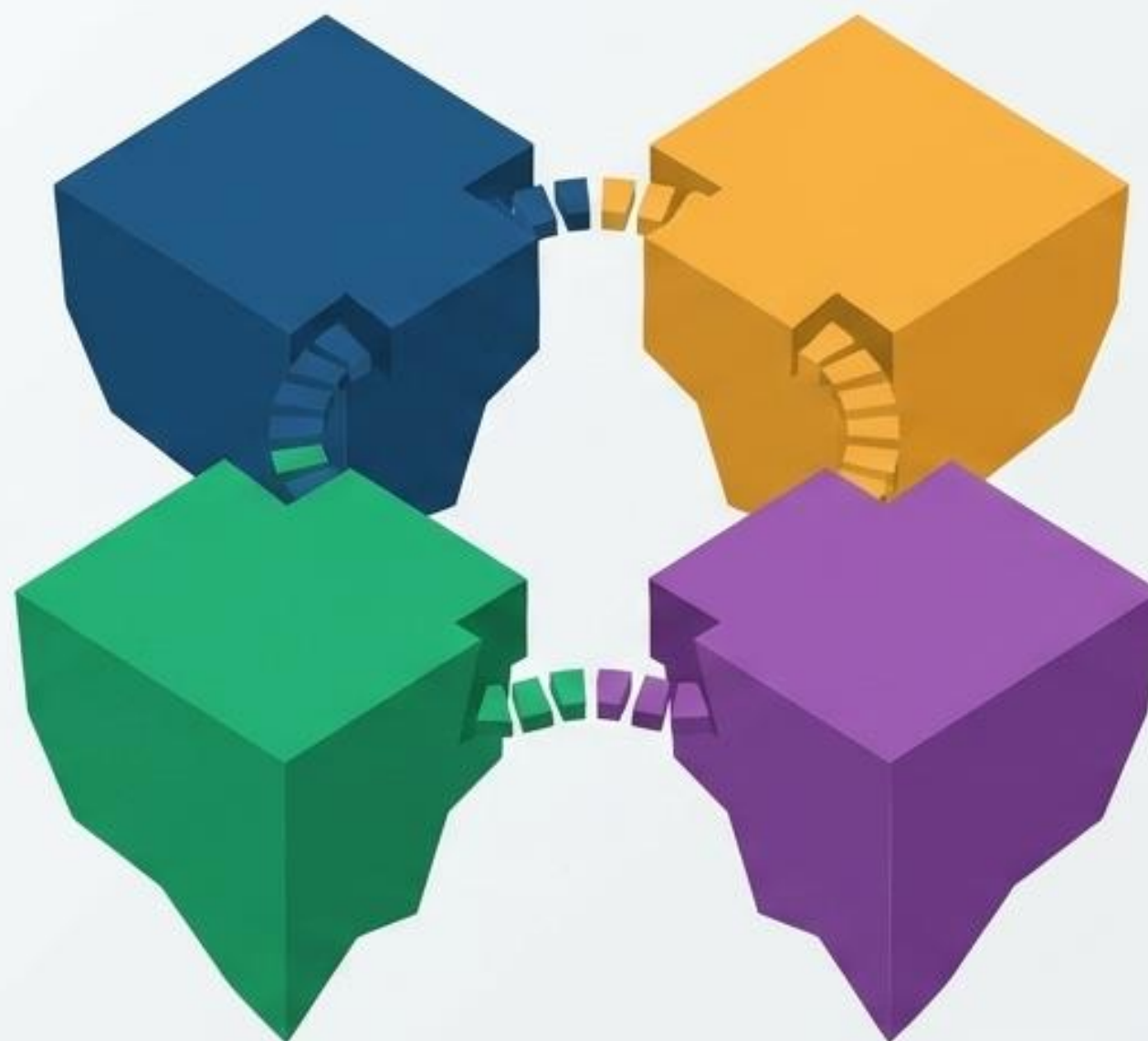
Focuses on digital preservation, events, and audit trails.

## Dublin Core

Focuses on lightweight standard resource discovery.

## Schema.org

Focuses on web visibility and SEO indexing.



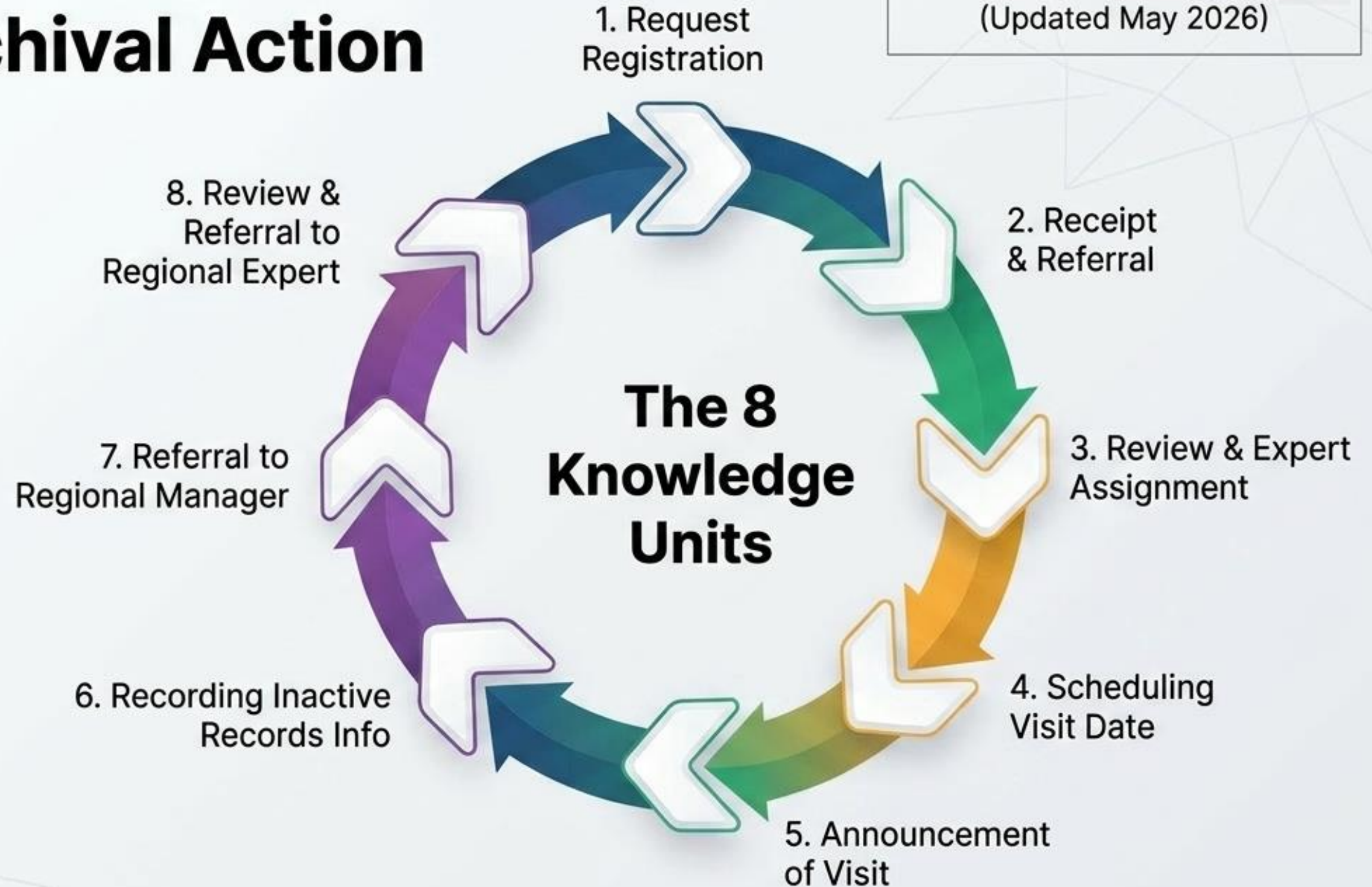
**Metadata Islands:** Independent design leads to data redundancy, semantic gaps, and the loss of relational context in Linked Data infrastructures.



# Reframing Standards Through Archival Action

Analyzing metadata standards not as static elements, but as dynamic tools responding to concrete archival activities.

Case Study: Appraisal Process of Inactive Records at NLAI  
(Updated May 2026)





# Bridging the Islands: Semi-Automated Semantic Alignment

## Lexical Similarity

Levenshtein distance & phonetic matching (resolving Persian script challenges).

## Semantic Similarity

Large Language Models (LLMs) utilized for concept embedding and description generation.

## Structural Similarity

Hierarchical position matching within ontologies.

AI Processing &

Candidate Generation

## Ex-Ante Expert Validation

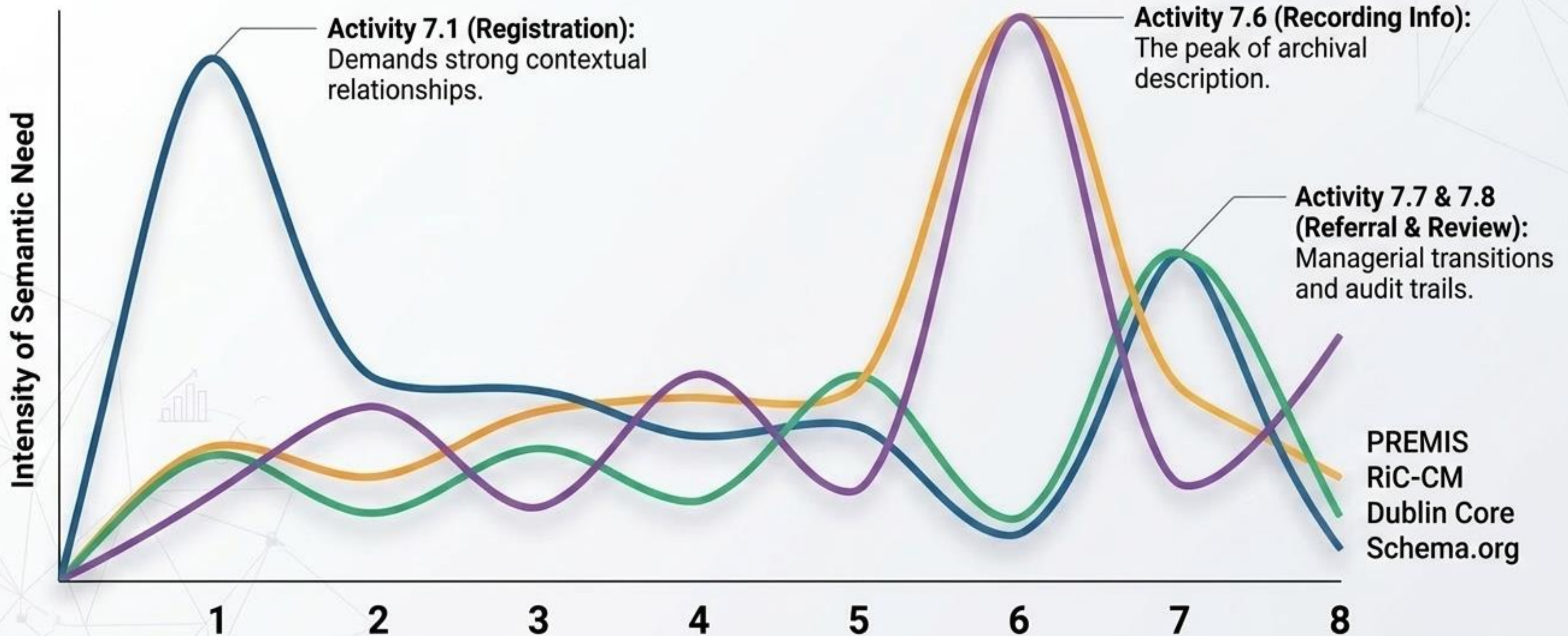
Panel of metadata experts ensuring cultural and semantic fidelity

**Validated Cross-Standard Alignments**



# Findings: The Dynamic Role of Standards

The role of each standard shifts dynamically depending on the specific archival activity. No single standard covers the entire lifecycle.





# Findings: Diagnostic Comparison Matrix

|                        | RiC-CM | Dublin Core | PREMIS | Schema.org |
|------------------------|--------|-------------|--------|------------|
| Organizational Context | ●      | ○           | ○      | ○          |
| Workflow Management    | ●      | ○           | ●      | ○          |
| Archival Description   | ●      | ●           | ○      | ◐          |
| Preservation/Audit     | ◐      | ○           | ●      | ○          |
| Web Exposure           | ○      | ○           | ○      | ●          |

**Takeaway:** Standards function as heterogeneous but inherently alignable Knowledge Organization Systems (KOS).



# Findings: Critical Gaps in Isolated Standards



## The Contextual Gap

DC and Schema.org fail to represent complex, multi-layered organizational and generational archival relationships.



## The Workflow Gap

DC lacks independent event modeling for critical managerial transitions and automated referrals.



## The Preservation Gap

RiC-CM and DC cannot document technical chain-of-custody or long-term digital preservation events.



## The Exposure Gap

RiC-CM, DC, and PREMIS are internally focused, remaining invisible to global search engine indexing.



# Proof of Concept: The Quantitative Impact

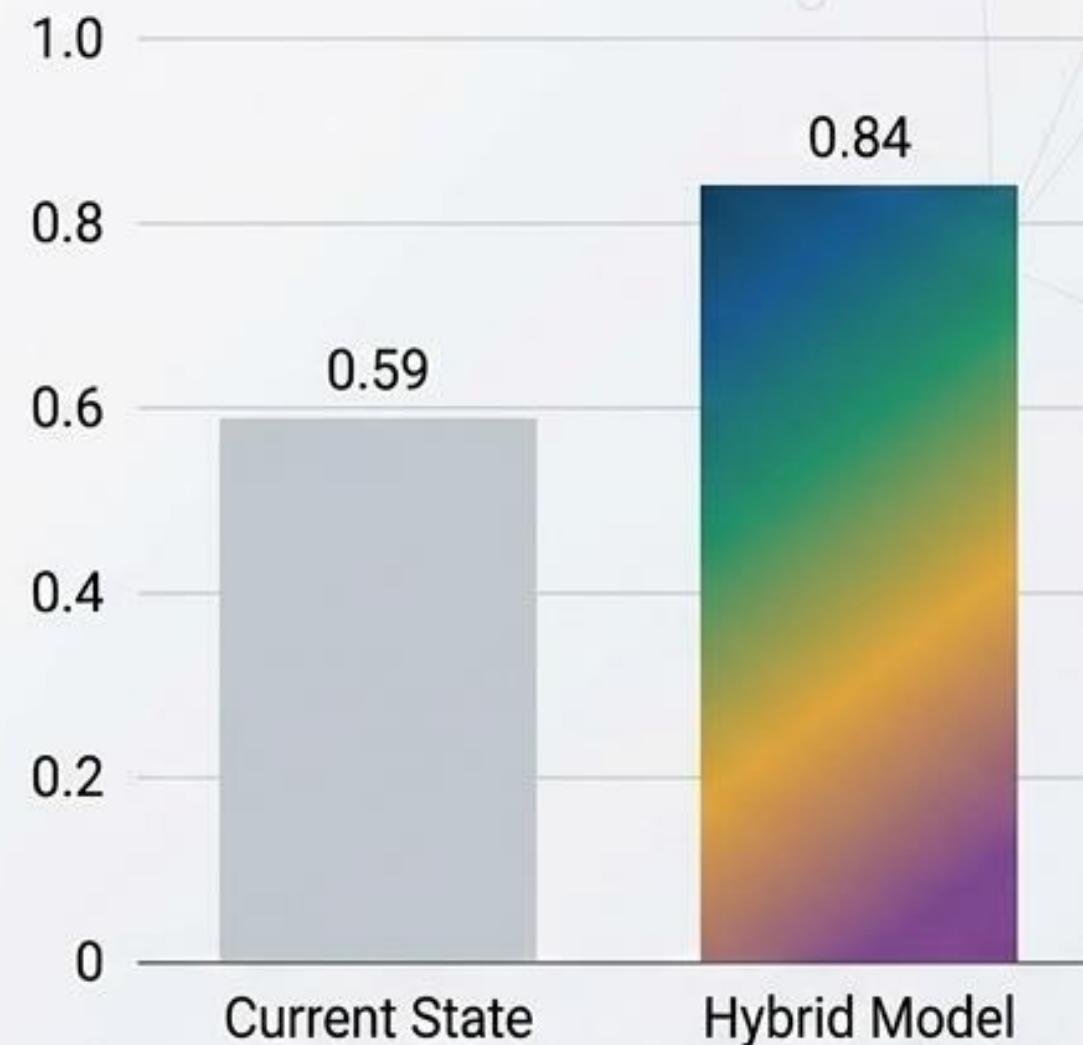
Pilot study testing 150 documents across 5 complex semantic search scenarios.

# +42.4%

Overall F1-Score Improvement (0.59 → 0.84)

Precision: +32.8%  
(0.64 → 0.85)

Recall: +53.7%  
(0.55 → 0.84)



**Highlight:** Scenario S3 — Combining document type, event, and preservation queries yielded a massive **66.7% improvement** in F1-Score over siloed systems.



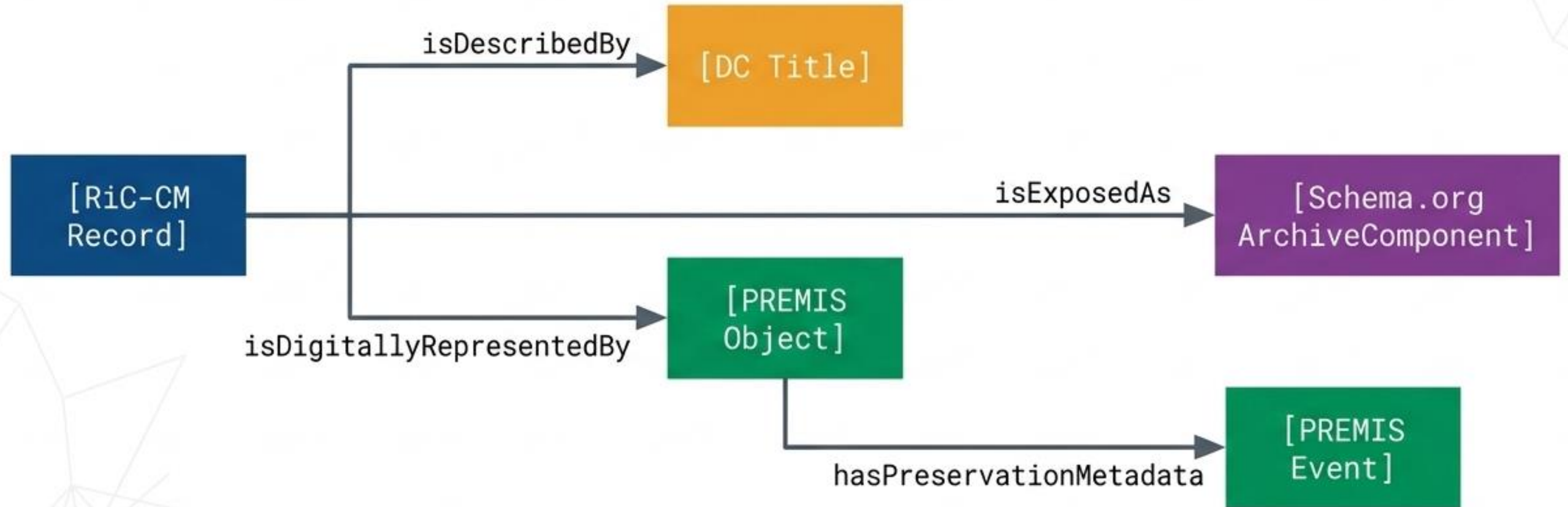
# The Proposed Model: 4-Layer Hybrid KOS Architecture





# Technical Implementation & Cross-Walking

OWL 2 Ontology Implementation designed for Linked Open Data (RDF Triples & SPARQL).



**Result:** Enables seamless logical reasoning and contextual navigation across all 4 layers simultaneously.



# Validation & Expert Consensus

Qualitative Assessment by a panel of 7  
**metadata and archival experts.**

**4.53 / 5.00**  
(Excellent)

Semantic Web Compliance: 4.86



Conceptual Coherence: 4.71



Extensibility: 4.71



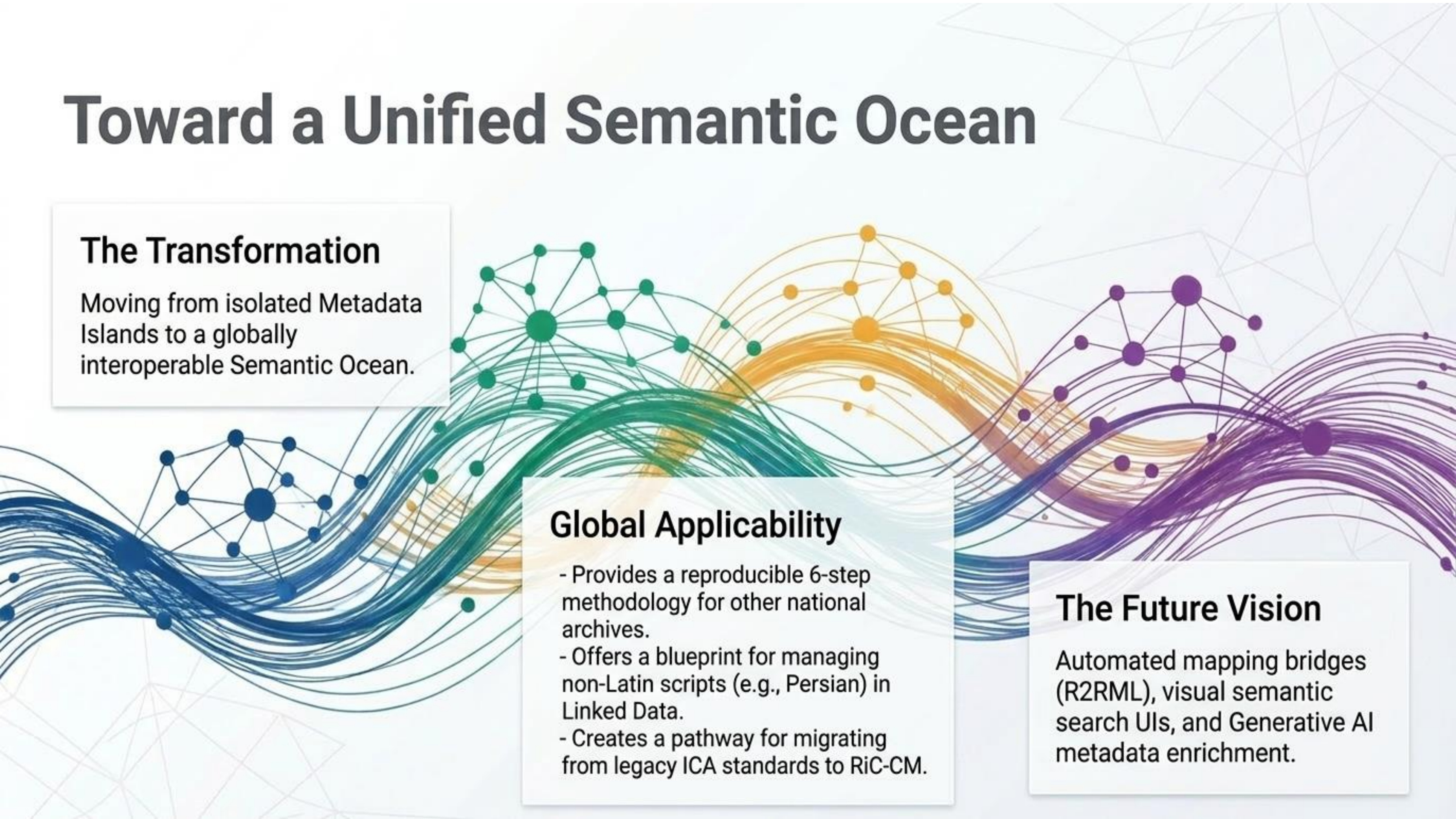
“

The simultaneous  
integration of standards  
and the concrete  
activity-based  
methodology provide a  
highly robust, culturally  
grounded framework.

”



# Toward a Unified Semantic Ocean



## The Transformation

Moving from isolated Metadata Islands to a globally interoperable Semantic Ocean.

## Global Applicability

- Provides a reproducible 6-step methodology for other national archives.
- Offers a blueprint for managing non-Latin scripts (e.g., Persian) in Linked Data.
- Creates a pathway for migrating from legacy ICA standards to RiC-CM.

## The Future Vision

Automated mapping bridges (R2RML), visual semantic search UIs, and Generative AI metadata enrichment.



# References

- Bountouri, L., & Gergatsoulis, M. (2011). The semantic mapping of archival metadata to the CIDOC CRM ontology. *Journal of Archival Organization*, 9(3–4), 174–207.
- Hawkins, A. (2022). Archives, linked data and the digital humanities. *Archival Science*, 22(3), 319–344.
- International Council on Archives (ICA). (2023). Records in Contexts – Conceptual Model (RiC-CM) Version 1.0.
- ISO. (2017). ISO 15836-1:2017 Information and documentation -- The Dublin Core metadata element set.
- Lourdi, I., Papatheodorou, C., & Doerr, M. (2009). Semantic integration of collection description. *D-Lib Magazine*, 15(7/8).
- PREMIS Editorial Committee. (2015). PREMIS Data Dictionary for Preservation Metadata, Version 3.0.
- Schema.org. ArchiveComponent & ArchiveOrganization.