DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Art of Tea - A Semantic Digital Library for the Interdisciplinary Research of Tea

Ilaria De Dominicis [1], **Regina Manyara** [1], and **Shiho Nakamura** [1]

[1] University of Bologna

Abstract

This paper proposes a methodology for creating small-scale semantic digital libraries using open standards, interoperable metadata, semantic technologies, and IIIF. While semantic digital libraries are often associated with large institutions, the proposed approach aims to make their development more accessible to researchers and small research groups. To demonstrate the methodology, we present Art of Tea, a work-in-progress semantic digital library dedicated to the interdisciplinary study of tea. Resources are described through standard ontologies, linked to external authority files, and organized through semantic relationships and Library of Congress Subject Headings. By combining accessible tools with established standards, the project illustrates how semantic digital libraries can be developed and adapted to different research domains

1. Introduction

In recent years, semantic technologies have increasingly shaped the way digital libraries are designed and accessed. While traditional digital libraries primarily make resources searchable, semantic digital libraries structure them through machine-readable metadata that encodes not only what objects are, but also how they relate to one another.

Building a semantic digital library (SDL) often requires technical expertise, specialized personnel, and established implementation frameworks [1]. As a result, fully realized SDLs are usually limited to large institutions such as national libraries and universities.

This project starts from a different premise: that building an SDL should not be an exclusive practice. Rather than adopting a pre-built platform, we propose a set of principles and best practices to guide researchers, students, and cultural heritage practitioners in creating interoperable, semantically structured collections from scratch.

To demonstrate and evaluate this methodology, we developed Art of Tea as a case study. Tea culture provides a rich and diverse domain through which to illustrate the different stages of the process, from material selection and metadata modelling to semantic enrichment and knowledge representation. The resulting library is therefore not only a standalone collection, but also a practical example of how the proposed methodology can be applied to other research topics and domains.

2. Methodology

Our methodology adapts the three-phase framework proposed by Tomasi [2]: Analysis of the Domain, Knowledge Organisation, and Knowledge Representation; prioritising tools and resources that are accessible, reproducible, and interoperable at each stage.

Student Forum

Available Aug 01, 2026

Correspondence to
Ilaria De Dominicis
ilaria.dedominicis2@studio.unibo.it

2.1. Analysis of Domain — Data Gathering and Open Sources

When building a digital library, the selection of materials depends on the goals of the project. Whenever possible, however, it is advisable to prioritize open-access and public-domain resources, as they facilitate reuse, ensure legal accessibility, and support the broader aim of making knowledge available to others, according to the FAIR principles [1]. The selection phase also shapes many later decisions including metadata design and the overall scope, and factors such as resource types, team size, and availability of descriptive information should be considered from the outset.

In our case study, we selected textual materials related to tea culture from publicly available sources such as the Internet Archive[1] and the Library of Congress[2], which provide access to books that are open access or in the public domain. However, relevant materials are not always openly licensed, and copyright information may sometimes be unclear or absent. In such cases, a useful approach is to contact the rights holder directly and request permission for use. In our case, for *The Classic of Tea*[3], we determined that the work held exceptional cultural and historical significance as the earliest known book on tea. Since the translation was not openly available for reuse, we contacted the translator directly and obtained permission for academic use.

Our experience also highlighted that resource selection should not be considered independently from metadata planning. Establishing clear criteria early helps ensure consistency and long-term sustainability, particularly for small teams.

2.2. Knowledge Organisation — Structuring the Metadata

For the organisation of semantic knowledge, we propose a two-level framework. The first level handles the structuring of the resource metadata, while the second level focuses on the structuring of the subject classification of the resources.

The description of the resource metadata can be handled by existing schemas such as BIBFRAME[4] and DCTERMS[5], which cover both literary and non-literary resources. For more extensive descriptions, links can be created to external authority files such as GeoNames[6] for locations and VIAF[7] for people. In our prototype for *Art of Tea,* we used a dedicated namespace, *ArTea*, to hold the metadata description of our resources.

For the description of the subject classification, we introduced a new subjecting hierarchy in the namespace *aspect*. While other established vocabularies like LOC and Dewey already provide extensive subjecting, these have been designed with large-scale libraries and other knowledge organisation projects in mind. What *aspect* provides is a

1
2
3
4
5
6
7

way for the individual researcher to build up their own subjecting hierarchy based on the study of their specific domain.

This structure contains four main classes: ResearchObject, Aspect, BroadArea, and NarrowSubject. We used CIDOC-CRM[8] to provide the semantic structure between the classes, defining ResearchObject as rdfs:subClassOf E73_Information_Object and the other three classes as rdfs:subClassOf E90_Symbolic_Object. Aspect identifies the scholarly domains which relate in some way to the topic at hand, loosely following the Dewey Decimal Classification[9]. BroadArea provides further specifications for each aspect leading to the finer specificity of NarrowSubject, which identifies specific topics of study. These topics can be identified and curated from all the resources used to create the online library. Individuals of NarrowSubject can be aligned with the Library of Congress Subject Headings (LCSH)[10], which provides a controlled vocabulary for subject headings and also allows the library to be interoperable with the wider LOC infrastructure.

By establishing these four new classes, we allow the researcher to extend subject knowledge incrementally, using rdfs:comment to add annotations and rdfs:seeAlso to link resources within the library and to external networks such as LCSH.

Figure 1 provides an example of how we adapted the *aspect* namespace to build up a view of the study of tea. The example shows how the book "Journey to the Tea Countries of China" can be broken down into several broad areas and narrow subjects which all form part of the knowledge of the aspect, tea. As illustrated in Figure 1, a single resource may be connected to multiple aspects such as History and Botany simultaneously, enabling users to approach the collection from different disciplinary perspectives. The full semantic structure used in the *Art of Tea* prototype can be found in RDF TTL format on the project's GitHub repository[11].

2.3. Knowledge Representation — Delivery and Presentation

For the representation of digital objects, we adopted the International Image Interoperability Framework (IIIF)[12] as the delivery and presentation layer of the library, complementing the semantic layer established through the *ArTea* graph and the *aspec*t classes. IIIF is well suited to small-scale digital libraries for three reasons: firstly, it enables the delivery of high-resolution visual content without dedicated server infrastructure. Secondly, its widespread adoption by major cultural heritage repositories including the Library of Congress and the Internet Archive ensures interoperability with existing collections [2], [3]. Thirdly, its manifests are JSON-LD documents that can carry descriptive metadata, making them a natural bridge between the presentation and semantic layers.

8

9

10

11

12

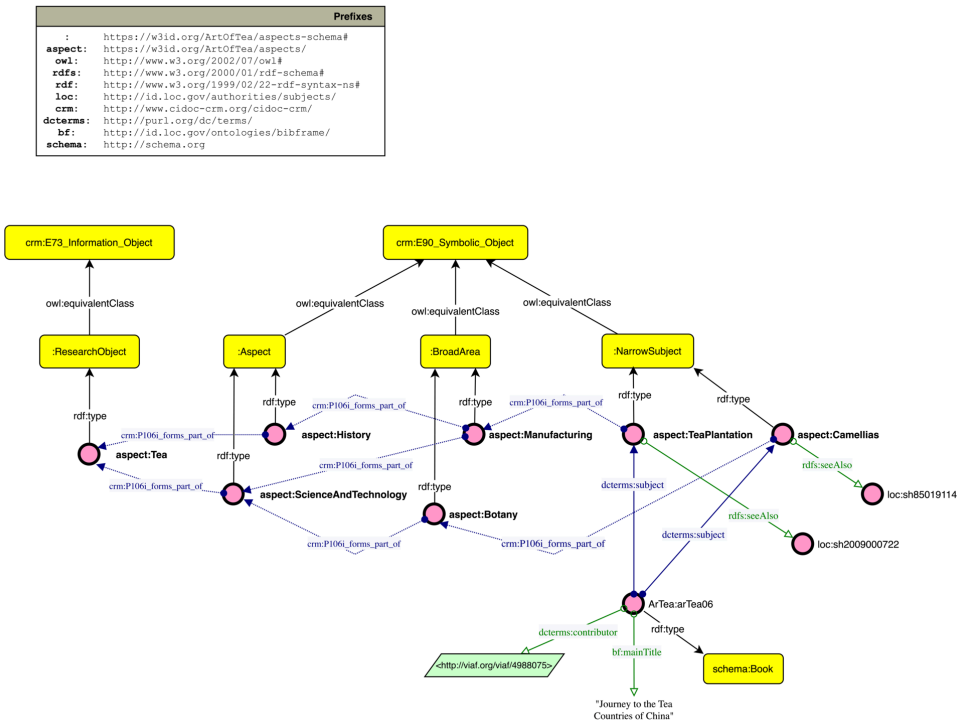

Figure 1. *Sample view of the* Art of Tea *semantic structure, using internal and external ontologies and standards to highlight different paths of research.*

This project uses the Image API for displaying artworks and the Presentation API for displaying digital text objects. Manifests were built using open tools[13] and enriched with descriptive metadata drawn from the *ArTea* graph; canvases and ranges express the hierarchical structure of chapters and sections, preserving the intellectual organisation of each work within the Presentation API specification. Where existing manifests were already provided by the Internet Archive, these were reused directly. Resources are presented via the Universal Viewer[14] embedded within the project website.[15]

3. Limitations and Future Work

As a work-in-progress project, *Art of Tea* remains limited in both scale and evaluation. The proposed methodology has so far been tested only through a single case study, and its application in other domains is yet to be explored. Furthermore, while the prototype demonstrates the technical feasibility of the approach, its usefulness for end users has not yet been formally assessed. Regarding the technical infrastructure, the current implementation relies on GitHub Pages[16] for hosting and IIIF-based tools for content delivery, prioritising flexibility in interface design. This approach is well suited to a small-scale prototype, but as a collection grows in size and complexity, established platforms such as Omeka, Islandora, or Greenstone[17] may offer significant advantages in

13

14

15

16

17

terms of content management, user administration, and long-term sustainability. Future work will focus on expanding the collection through IIIF annotations and Search API integration, applying the methodology to additional domains, and gathering feedback from researchers and students to further refine the framework.

4. Conclusion

In his groundbreaking essay *As We May Think*, Dr Vannevar Bush describes the Memex, "a future device for individual use, a sort of mechanized private file and library" [4], almost a century before the possibility for such a device could exist. Now, thanks to semantic technologies, we find ourselves at such a time.

Our primary goal was to demonstrate that building a semantic digital library is not limited to large institutions, but can instead be approached through accessible tools, shared standards, and clear methodologies. In order to do so, the *Art of Tea* prototype was born, and the project is an ongoing proof of concept.

Making semantic tools useful to the individual researcher brings us closer to the device Bush envisioned. His personal digital library was guided by the associative nature of the human mind. The flexibility of the subject descriptions in our digital library allows us to enrich our research collection with interdisciplinary knowledge as well as how we may think.

Acknowledgements

We would like to express our gratitude to our supervisor, Professor Giovanni Colavizza, for guidance and encouragement throughout this project.

References

[1] K. Shahzad and S. A. Khan, “Factors affecting the adoption of integrated semantic digital libraries (SDLs): A systematic review,” *Library Hi Tech*, vol. 41, no. 2, pp. 386–412, 2023, doi: 10.1108/LHT-05-2022-0231.

[2] F. Tomasi, “Digital humanities e organizzazione della conoscenza: Una pratica di insegnamento nel LODLAM,” *AIB Studi*, vol. 60, no. 2, pp. 411–425, 2020, doi: 10.2426/aibstudi-12068.

[3] S. Snydman, R. Sanderson, and T. Cramer, *The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images*. in Proceedings of IS&T Archiving 2015. 2025, pp. 16–21. doi: 10.2352/issn.2168-3204.2015.12.1.art00005.

[4] V. Bush, “As we may think.” [Online]. Available: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

[5] N. Freire, G. Robson, and J. B. Howard, “Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org,” *International Journal on Digital Libraries*, vol. 21, pp. 19–30, 2020, doi: 10.1007/s00799-018-0259-5.

[6] M. Wilkinson, M. Dumontier, and I. Aalbersberg, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, 2016, doi: 10.1038/sdata.2016.18.