# A Linked Open Data Infrastructure for Promoting the Educational Use of Digital Archives

Masao Takaku[1] [0000-0002-2458-6988], Yuka Egusa[2] [0000-0002-5475-1657], Satoshi Enomoto[3], Masao Oi[4] [0000-0002-6750-5092], Yumiko Ariyama[5], and Takayuki Ako[6] [0000-0001-7682-3036]

[1] University of Tsukuba, Kasuga 1-2, Tsukuba, Ibaraki, Japan
masao@slis.tsukuba.ac.jp
[2] National Institute for Educational Policy Research, Kasumigaseki 3-2-2, Chiyoda, Tokyo, Japan
[3] Japan Women's University, Mejirodai 2-8-1, Bunkyo, Tokyo, Japan
[4] Doshisha University, Karasuma-higashi-iru, Imadegawa-dori, Kamigyo, Kyoto, Japan
[5] Shiga Bunkyo Junior College, Tamura-cho 335, Nagahama, Shiga, Japan
[6] Tokyo National Museum, Ueno Park 13-9, Taito, Tokyo, Japan

**Abstract.** Digital cultural heritage is increasingly available from memory institutions. However, its integration into formal education remains limited due to a lack of education-specific metadata and weak alignment with curricula. This paper addresses this gap in the Japanese context, where national curricula are centrally defined but rarely linked to digital resources. We present a Linked Open Data (LOD) infrastructure that connects curriculum standards (Course of Study), textbook units, and digital archive content through a reusable, semantically modeled framework. We contribute to implementing a scalable, curriculum-aligned metadata foundation that supports practical educational use. We describe the structure, vocabulary design, and implementation of the infrastructure, and demonstrate its utility through linked applications such as search tools and visualizations. This work lays the groundwork for connecting curricula with cultural resources, enabling more effective discovery, reuse, and integration of archival content in educational settings.

**Keywords:** Digital archives · digital cultural heritage · educational use · Linked Open Data (LOD) · metadata

## 1 Introduction

Digital archive services are increasingly widespread as libraries, museums, archives, and other memory institutions digitize cultural resources and make them accessible online. In Japan, the government officially launched a search portal site for digital archives called Japan Search in 2020 [1]. Users can search, discover, and use tens of millions of materials through such services. This paper uses the term "digital cultural heritage" to refer broadly to digitized cultural assets held by memory institutions, including historical documents, artworks, and artifacts. We refer to the systems and services that collect, store, and provide access to such resources as "digital archives." For consistency, we use the term "digital archives" throughout this paper to represent both the resources and their associated systems.

The educational use of digital archives is gaining attention as a promising application [42]. Using digitized primary sources in schools can deepen learning by encouraging critical thinking beyond textbook information. Furthermore, memory institutions can enhance their societal role by incorporating digital archives into educational settings. Such applications can justify digitization costs and strengthen the role of memory institutions in digital education.

However, the use of digital archives in educational settings remains limited. Various institutions provide digital archives in a decentralized manner, making them difficult to search, access, and use effectively. While several studies [42][33][44] have discussed strategies for incorporating digital cultural heritage into classroom settings through lesson materials and activities, these efforts remain fragmented and lack a unifying infrastructure to support sustainable integration across educational contexts.

In order to enhance the educational value of digital archives, it is essential to assign appropriate metadata and establish connections between archive materials and educational content [33]. Existing metadata in digital archives is often limited to basic content descriptions. It does not support direct links to curricula, learning activities, or pedagogical goals [33]. Addressing this gap requires a structured and interoperable framework to describe, connect, and reuse resources in a way that aligns with educational practice.

Many projects have employed Linked Open Data (LOD) to enrich and share metadata for digital archives [18][9][32][22][29]. However, most of these initiatives focus on academic or general public use. Only a few target the specific needs of the educational domain, and little attention has been paid to building reusable, domain-specific infrastructure tailored to educational content and practices.

This study focuses on designing and implementing an LOD-based infrastructure that enables the integration of digital archives into educational contexts. Building on tools such as Textbook LOD [3][28][27], Teaching Unit LOD, and Course of Study LOD [2], we construct a unified data infrastructure that connects learning resources and digital archive materials using LOD technologies. Our work centers on developing a reusable and extensible platform that supports linking digital archives with educational content. We analyze the structure of this infrastructure and demonstrate how it can serve as a bridge between cultural heritage resources and real-world teaching practices.

Our contribution lies in designing and operationalizing a practical LOD-based infrastructure that directly responds to the unmet needs of educators and learners, rather than proposing a novel technical method. Through close collaboration and discussions among the co-authors, two of whom have extensive experience in educational practice, we identified core challenges such as the difficulty of discovering educationally relevant archive content, the lack of alignment with curriculum objectives, and the limited metadata provided by cultural institutions. The proposed infrastructure addresses these through reusable, semantically connected components (e.g., textbook units, curriculum items) that support educational discovery and instructional planning. The significance of this study lies in developing a bridging architecture that facilitates the integration of digital archives into authentic educational contexts.

## 2 Related Works

### 2.1 Use of Digital Archives in Education

Digital archives provide broader access to primary and historical materials, offering new opportunities for teaching and learning [12]. Since the 1990s, projects like the Library of Congress's American Memory [16, 17, 20, 51, 14] have demonstrated their educational potential.

In recent years, initiatives such as Europeana's Teaching with Europeana program [6], Maekawa's local learning platform [34], and the S×UKILAM project [41, 43] have promoted the use of archival materials in educational settings.

Despite these efforts, challenges persist. Teachers and learners struggle to find suitable materials due to dispersed archives and insufficient metadata [12]. Moreover, the weak links between archival content and curricula hinder integration into formal education [33, 15]. Efforts to address this include adding curriculum-aligned metadata to archival resources [33, 24, 45].

Vlachou et al. [52] further highlight the need for interoperable metadata models integrating education, cultural heritage, and XR technologies, underscoring metadata's key role in linking digital content to learning.

To address these challenges, we have built an "education-oriented LOD infrastructure" that structures curriculum elements as LOD and semantically connects them to archival materials, supporting more effective educational use.

### 2.2 Metadata for Education

Describing learning resources with appropriate metadata has long been studied in education and information science. Such metadata provides bibliographic details and pedagogical attributes (e.g., age group, difficulty level, context, curriculum alignment). For instance, the IEEE Learning Object Metadata (LOM) standard [11] defines items covering learning context, user role (e.g., teacher, student), difficulty, and age range [19]. Standards like LOM promote consistent metadata attachment, improving discoverability and interoperability across learning management systems.

To further support this, the U.S. developed the Achievement Standards Network (ASN), a machine-readable RDF model of K-12 curriculum standards [46]. ASN enables the alignment of resources with learning objectives (e.g., matching a history lesson to a WWII standard), enhancing the retrieval of standards-based materials. Tavakoli et al. [48] show that metadata can predict the quality of Open Educational Resources (OERs) with high accuracy using machine learning, suggesting metadata's utility in quality assessment as well as discovery and alignment.

In Japan, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) provides national curriculum guidelines, the Course of Study (CoS), and assigns 16-digit CoS codes to each learning objective [36]. These codes offer identifiers at fine-grained levels of content.

Recent studies have explored the use of structured data in education. Abu-Salih et al. [13] provide a literature review on knowledge graphs (KGs) in education, highlighting their roles in self-directed learning, analytics, curriculum modeling, and semantic content representation. They observe a shift from static, descriptive metadata toward dynamic, semantically enriched structures linking learners, content, and pedagogical goals. They emphasize the growing use of RDF-based models and ontologies for aligning resources with standards and learner profiles. This broader trend is exemplified by international projects such as ASN [46] and Europeana's Teaching with Europeana [6], which demonstrate the value of using metadata and RDF frameworks to connect educational resources with curricula and cultural heritage.

This study extends that semantic approach to the Japanese context by leveraging the nationally standardized Course of Study (CoS) codes. These codes provide a fine-grained, legally defined framework for describing curriculum content. We utilize them to interlink curriculum guidelines, textbook bibliographic data, and unit-level instructional content. While existing projects often focus on broader standards or cultural content, our approach is characterized by its fine-grained alignment of textbook units with the CoS codes. Through this structure, we aim to bridge formal curricula and culturally meaningful learning resources in a reusable and interoperable way.

## 3 Methods

### 3.1 Overview

In this study, we designed and developed a LOD infrastructure to serve as a practical bridge between curriculum standards, textbook content, and digital archive materials. Rather than introducing new technologies or algorithms, we focused on developing a practical and sustainable infrastructure informed by real-world educational needs.

To this end, we adopted a layered modeling strategy that reflects the structure of the Japanese school curriculum. The platform comprises three interrelated datasets: the Course of Study (CoS) LOD, which formalizes government-issued learning objectives; the Textbook LOD, which captures bibliographic metadata of authorized textbooks; and the Teaching Unit LOD, which models chapter-level units from textbooks. These layers are linked through consistent URIs and shared properties, enabling semantic traversal across curriculum objectives, textbook content, and related teaching resources.

The need for interoperability, extensibility, and clarity for human users and machines guided each design choice. We adopted standard vocabularies such as Schema.org [31] and DCMI Metadata Terms [23] to describe core metadata. URI design followed consistent patterns based on existing identifiers, such as the official CoS code. We used the Shapes Constraint Language (SHACL) [10] to validate data consistency, document data shapes, and express metadata application profiles during the development. These choices support robust interlinking among the datasets and promote reliable reuse in educational applications.

This chapter describes the components of this educational LOD infrastructure, URI design, vocabulary design, and resource linking. We also explain how applications can be integrated using the infrastructure.

### 3.2 Components and Designs of LOD infrastructure

This section outlines the components and design principles of the proposed LOD infrastructure. This infrastructure integrates multiple education-related information resources as LODs. It converts them into an interconnectable structure, thereby improving the discoverability, scalability, and reusability of the data for educational purposes.

**Datasets that make up the LOD infrastructure** The proposed LOD infrastructure is composed of the following three LOD datasets:

1. Course of Study LOD (CoS LOD)
   This dataset was created by reorganizing the "Course of Study Code" [36] provided by the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) in 2020 in spreadsheet format such as Excel and extracting the structure and content of CoS established by MEXT, including its hierarchical structure (subject → grade → objectives → content), and creating an RDF data model. Each item is assigned a unique URI based on the CoS Code, and hierarchical relationships and links to related content are described using vocabulary such as Schema.org [5].
2. Textbook LOD
   Based on the bibliographic information created by the Education Library at the National Institute for Educational Policy Research for library services, this is an RDF dataset of bibliographic information metadata. In addition to basic bibliographic information such as the title and editor of the textbook, this dataset contains education-specific information such as the subject, year of textbook authorization, and year of use.
3. Teaching Unit LOD
   This LOD dataset defines educational content at the unit level based on a textbook's chapter and section structure. The dataset is provided with the same namespace URI and on the same site as the textbook LOD dataset mentioned above. It is structured by linking textbook unit information listed in the compiling prospectus for the textbooks [37] provided by the textbook publisher at the time of screening with a CoS resource that indicates the content to be learned in that unit.

**URI design** Each LOD resource is assigned a persistent and identifiable URI using the following structure:

1. CoS LOD: Base URI is ⟨https://w3id.org/jp-cos/⟩.
   Individual resource URIs are expressed in the following format, with the 16-digit CoS Code added to the base URI: https://w3id.org/jp-cos/8220263213000000
2. Textbook LOD: Base URI is ⟨https://w3id.org/jp-textbook/⟩.
   Individual resource URIs are expressed as follows, with the school type, examination year, textbook symbol, and textbook number connected with a slash symbol: https://w3id.org/jp-textbook/小学校/2019/社会/602
3. Teaching Unit LOD: Base URI is ⟨https://w3id.org/jp-textbook/⟩ (same as textbook LOD).
   Individual resource URIs are expressed in the following format using a 14-digit sequential management number: https://w3id.org/jp-textbook/AA100000002045

Please note that our infrastructure, especially in the Textbook LOD, follows the IRI (Internationalized Resource Identifier) specification [25] rather than being limited to ASCII-based URIs. While many LOD projects adopt ASCII-only URIs as a conventional and convenient practice to support international use, our design intentionally allows Japanese characters in IRIs to reflect authentic educational contexts and provide more intuitive identifiers for local users. As IRIs natively support non-ASCII characters, this approach remains technically valid and interoperable, while also promoting a broader view of internationalization that embraces diverse writing systems.

**Vocabulary Design** In constructing the LOD infrastructure, we ensure interoperability with other LOD projects and external services by adopting a standard, widely used, and highly versatile vocabulary. Table 1 shows the vocabulary used in designing the LOD infrastructure. Schema.org [5] and DCMI Metadata Terms [23] are used to describe basic metadata and bibliographic information of resources. At the same time, RDF Schema [21] and SHACL [10] are used to describe and define metadata schemas. However, since standard vocabulary is insufficient to express educational terms and concepts in the educational system, we independently defined a new vocabulary based on the base URI of the dataset itself.

For example, properties for concepts such as the following are relevant:

- `textbook:grade`: Target grade of a textbook or teaching unit.
- `textbook:authorizedYear`: Authorization year of a textbook.
- `textbook:subject`: Subject that the resource corresponds to.
- `cs:cosItem`: CoS item that the resource corresponds to.
- `cs:sectionNumber`: Section number of the description of a CoS item (e.g., A and (1)).

A key principle in designing a LOD infrastructure is to ensure that resources are interconnected, not isolated. For example, we aim to create a structure that

Table 1: List of vocabulary used in the LOD platform (excerpt)

| Vocabulary name | Main purpose/role | Examples of properties used |
|---|---|---|
| Schema.org [5] | Description of bibliographic information metadata for textbooks and units, description of hierarchical structure of curriculum items | schema:name, schema:publisher, schema:isbn, schema:description, schema:isPartOf |
| DCMI Metadata Terms [23] | Description of basic resource information (identifier, source, etc.) | dct:identifier, dct:source |
| RDF Schema (RDFS) [21] | Class and property definitions, descriptions of basic type relationships between resources, reference descriptions of related web pages | rdfs:label, rdfs:subClassOf, rdfs:seeAlso |
| SHACL (Shapes Constraint Language) [10] | Describing application profiles for LOD resources, description of constraints for consistency verification (validation) | sh:property, sh:path, sh:minCount |
| Original (our own) vocabulary (textbook:, cs:) | Expression of concepts specific to the education system, such as target grade, examination year, corresponding subjects, CoS items | textbook:grade, textbook:authorizedYear, cs:cosItem, cs:sectionNumber |

enables searches to begin with a CoS item associated with a teaching unit in a textbook and then to discover corresponding digital archive materials and external teaching materials sequentially. To achieve this, we proactively add cross-reference links (schema:isPartOf, rdfs:seeAlso, etc.) to each resource to enhance the navigability of the LOD.

In particular, among the three LOD datasets, the CoS LOD is a hub that connects various teaching materials and content. Figure 1 demonstrates links to a CoS LOD resource representing standardized learning content. Not only are links made from a teaching unit resource in a specific textbook, but also from teaching material content of the S×UKILAM workshop [41] and digital content in the Tokyo Gakugei University Library Digital Collection [50]. These links are connected using the LOD property cs:cosItem we provide. This way, not only can unit resources within the LOD platform be linked to the CoS LOD resources, but teaching materials and other digital content can also be linked to the CoS LOD resources. Because these links are distributed and loosely coupled in a LOD way, external teaching material providers can freely add links, which is advantageous. Even if an external educational material provider does not support LOD links, if the external content has a CoS code number in a metadata record, it is possible to use a web API to search for it and connect content that shares the CoS code. Figure 1 also shows links to educational video material provided by the public broadcaster NHK [40].

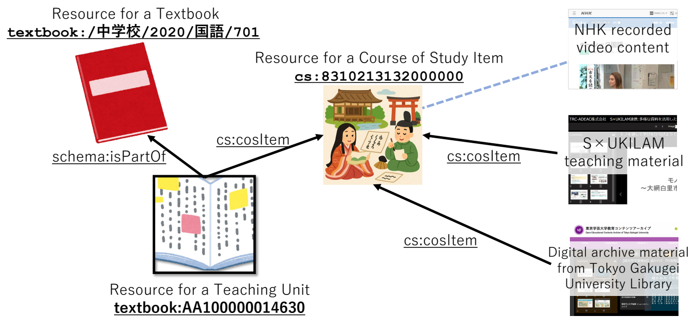


Fig. 1: Links to a course of study item resource from the LOD datasets and from digital archive contents and external educational materials

### 3.3 Designing for Data Updates and Expansion

The three LOD-based datasets originate from different information sources. CoS guidelines have undergone revisions approximately every ten years. Each CoS version uses a different code system, which also applies to the source data for the CoS LOD. When creating a LOD dataset, the process treats the CoS codes directly as LOD resources, meaning that the items in each revised version appear as distinct LOD resources. In addition, each CoS version serves as a basic LOD resource, linked to its corresponding items, thereby clarifying which version each set of items belongs to.

Future revisions of the CoS will likely introduce new codes, and the corresponding LOD resources will be added accordingly. Because of this structure, updatability and scalability are not major concerns. However, a challenge arises when aligning learning resources linked to older CoS with the latest versions. To address this, it will be necessary to re-associate existing learning resources with the new CoS items. Ideally, this process would involve defining explicit relationships between old and new items by preparing mapping data that captures their correspondences. Managing these relationships between different versions will remain important for future development. Though not yet implemented, mapping between old and new Course of Study versions is important. In Japan, curricula are distributed without alignment annotations, complicating item-level links. Nonetheless, partial alignment is feasible in structured subjects like science or math. We plan to explore using SKOS [38] to support interoperable, machine-readable mappings.

As mentioned in Section 3.2, textbook and teaching unit data are based on two information sources, the Education Library and the MEXT. We obtain the data and build and update the LOD dataset as soon as these information sources are updated. However, the textbook compiling prospectus [37] provided by the MEXT in PDF format, which is in a human-readable booklet format but not machine-readable. As a result, updating the LOD-based dataset based on this information requires labor-intensive manual review and transcription, which puts a significant constraint on updateability. In our experience, mapping approximately 300 textbooks required four months and cost around USD 6,000. We advocate for curricula released in machine-readable formats (e.g., Excel, JSON) to scale, reducing integration costs. We are engaging stakeholders (e.g., government, publishers, educators) to discuss cost-sharing and maintenance strategies.

### 3.4 Approach to Data Quality Assurance

We use two validation tools to ensure the quality of the LOD infrastructure.

The first tool is an in-house script that checks for disconnected LOD resources. For example, librarians at the Education Library manually transcribe textbook property values such as subjects and grades. Although rare, transcription errors can cause broken links between resources. The script performs isolation checks to identify such issues easily.

The second is validation based on SHACL [10], which specifies constraints on the graph shapes of LOD resources. For example, a resource of the cs: TeachingUnit class, which represents a teaching unit, must have a schema:name property that corresponds to the unit name, the value of the textbook:grade property, which indicates the target grade, is a numeric expression (xsd:integer), and a maximum of six links can be added per unit resource, and these constraints are specified and described in SHACL. These SHACL constraint definitions facilitate the detection of errors in class assignments and property occurrences within the dataset. We use SHACL not only for constraints but also for documentation, essentially organizing descriptions and communicating them to users as metadata application profiles [35].

### 3.5 Applications

The proposed LOD platform is designed to support practical educational use by integrating integration visualization and search applications. Informed by collaboration with teachers, librarians, and educational content developers, we identified three key user profiles:

**Teachers and educators (E):** When preparing classes, they search for external teaching materials and resources corresponding to textbook units and curriculum guidelines .

**Educational materials developers (D):** They design educational content while ensuring alignment with CoS guidelines.

**Learners (L):** They search for learning materials related to specific teaching units and topics for self-directed study.

These users require an intuitive interface that allows them to browse and explore the relationships among the information stored in the LOD infrastructure and related teaching materials.

We envision the following applications, each tailored to the aforementioned user profiles.

**Curriculum Visualization Tool (E, D, L):** An application that visualizes the hierarchical structure of the Course of Study (subject → grade → objectives → content), allowing users to browse curriculum items and see linked textbook units and external resources.

**Federated Search Interface (E, L):** A search platform that integrates multiple LOD datasets, including textbook metadata, teaching units, and archive-linked materials, using a combination of SPARQL queries and full-text search. Users can filter by subject, school level, or curriculum item to locate relevant resources for both teaching and self-directed learning.

**Resource Recommendation Function (E, L):** Based on a selected CoS item or textbook unit, this component recommends linked digital archive materials or lesson content. Semantic relationships in the LOD graph allow dynamic generation of related materials, thereby improving discoverability.

**URI Reference and External Linking Function (E, D, L):** Each resource detail view displays its LOD URI and enables linking to external LOD resources such as Wikidata [8] or the National Diet Library Search [39].

While some of these application components are under active development, early prototypes have demonstrated the feasibility and usefulness of the LOD-based approach in real educational scenarios. The tools also serve as reference implementations for how structured curriculum-aligned metadata can enhance discoverability, integration, and reuse of diverse educational resources.

## 4 Results and Discussion

### 4.1 Basic Statistics of LOD Infrastructure

Textbook LOD includes textbooks for elementary, junior high, and high schools, covering Course of Study (CoS) revisions from 1989 to the current. As of July 2025, the dataset contains information on 9,178 textbooks, comprising 344,962 RDF triples.

Teaching Unit LOD is based on certified textbooks aligned with the current CoS (revised in 2017-2019). It includes unit-level information from all elementary and junior high school textbooks and about half of high school textbooks [49]. In the past year (2024–2025), we fully included unit-level data for all elementary and junior high school textbooks and expanded coverage for high school textbooks. This enhancement enables comprehensive navigation from textbooks to curriculum objectives, which was not feasible in earlier iterations of the dataset. The current dataset includes 37,082 units and provides 1,092,195 triples.

CoS LOD focuses on the two most recent CoS revisions. It includes CoS for kindergarten, elementary, junior high, high school, and special needs education schools. It represents all CoS-coded items as LOD resources. There are 39,958 such items, with 949,284 RDF triples.

Regarding the interlinking of datasets, 298,532 links connect unit resources to CoS items, with an average of 8.05 links per unit. Rather than simply transcribing CoS item references from the textbook compiling prospectus, the relationships were expanded to include hierarchically related items—both broader and narrower—based on the chapter and section structure. These related items were grouped per unit, enabling detailed semantic connections that are useful in various educational applications.

Additionally, the LOD platform links basic teaching materials and video content from NHK for School [40]. This web-based service incorporates CoS codes and offers search and display functionalities for teaching materials developed through the S×UKILAM workshop [41], transforming digital archive content into educational resources.

All components of the LOD platform — Textbook LOD, Teaching Unit LOD, and CoS LOD — use persistent URI identifiers assigned via w3id.org [53], managed by the W3C Permanent Identifier Community Group. The datasets are published using GitHub Pages [30], and the entire LOD infrastructure site is built with the static site generator `ttl2html` [47].

### 4.2 Application Prototypes

The LOD provider site visually presents its relationships through explicit linking to demonstrate the interconnectivity of the three LOD datasets. For example, as shown in Figure 2a, textbook-unit resources corresponding to CoS items are displayed as linked entries. Similarly, Figure 2b shows an example of a CoS resource linked to a list of related video materials from NHK for School, illustrating connections to external learning resources.

These link relationships can be easily implemented using publicly available Web APIs or the URI resolution functionality of LOD resources. For example, machine-readable data in RDF/Turtle format for any specific LOD resource can be retrieved by sending an HTTP request specifying that format. In contrast, external services such as NHK for School do not support LOD-based access. Instead, search results can be retrieved by submitting the relevant CoS code as a query to their dedicated Web API (e.g., the NHK for School API [4]). By combining such APIs with a client-side scripting language like JavaScript, link-based browsing and search functionalities like those illustrated in Figures 2a and 2b — can be readily implemented.

In addition, we have developed and released a "Textbook, Unit, and CoS Item Relationship Visualization Tool" [7], which visualizes the interconnections among datasets comprising the LOD platform. As shown in Figure 3, the tool represents direct link relationships between textbooks, teaching units, and CoS items as a network structure. In the visualization, blue nodes represent textbook resources, green nodes indicate teaching unit resources, and red nodes correspond to CoS item resources. Users can interactively explore the network by dragging and dropping nodes, helping them grasp the overall structure and interrelations among these resources. We developed this prototype application using the SPARQL engine, which enables flexible queries across the linked datasets and supports dynamic integration of CoS items, textbooks, and external teaching resources. We have made an experimental SPARQL endpoint available to support data reuse by external researchers and developers.


(a) Example of searching and displaying related relationships from course of study resources to textbook unit resources

(b) Example of search results for external teaching materials (NHK for School) from course of study resources using curriculum code

Fig. 2: Links and search capabilities on the LOD website (Source: [26])

Figure 3 illustrates the relationships between two elementary school history textbooks published by Tokyo Shoseki Co., Ltd. It highlights changes in the connections between teaching units and CoS items before and after textbook revisions. By analyzing these differences, teachers can compare editions and identify textbook variations, supporting classroom instruction and enhancing their understanding of textbook characteristics. This tool thus demonstrates the potential utility of LOD-based educational infrastructure in practical teaching scenarios.

Implementing the applications described in Section 3.5 can be achieved as extensions of the functionalities outlined above. However, these application programs are still under development and represent ongoing challenges for future work.

## 5 Conclusion

This study presented a Linked Open Data (LOD) infrastructure that connects curriculum standards, textbook units, and cultural heritage content to support the educational reuse of digital archives. Instead of proposing a novel algorithm or technical method, our contribution lies in designing and implementing a practical, curriculum-aligned metadata foundation. This foundation addresses long-

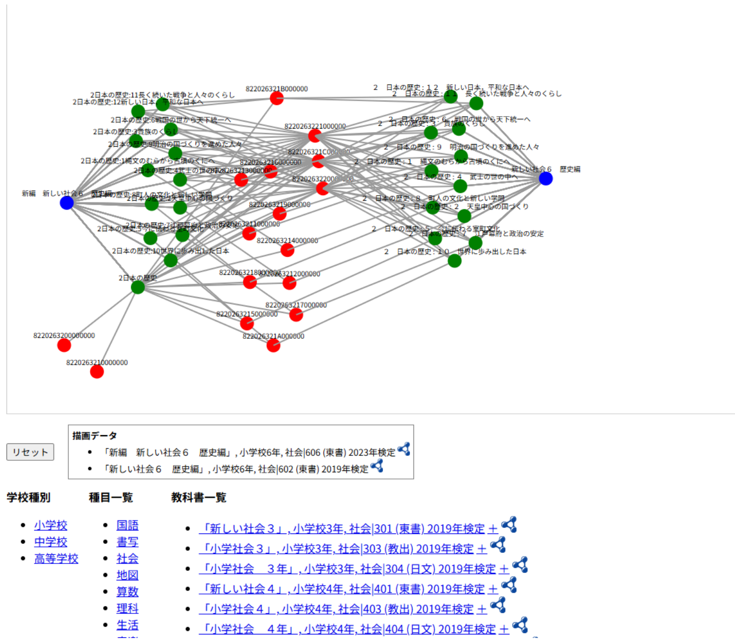


Fig. 3: Screenshot of the visualization tool for the relationships between textbooks, units, and CoS resource

standing challenges in the educational use of digital content, such as fragmented access, insufficient metadata, and a lack of alignment with learning objectives.

The infrastructure comprises three interoperable datasets — Course of Study LOD, Textbook LOD, and Teaching Unit LOD — linked through persistent URIs and shared vocabularies. With the recent completion of unit-level data for elementary and junior high school textbooks, the platform enables seamless navigation from curriculum goals to textbook units and linked learning materials, including digital archives and teacher-developed resources. This capability supports new applications, including curriculum-based resource discovery, personalized learning, and cross-institutional resource sharing.

Several issues remain for future work. First, further applications should be implemented, especially tools that connect and search across educational resources and digital archive materials. While a prototype network visualization tool has already been developed as shown in Section 4.2, additional and more diverse applications will be important. These include not only advanced search-oriented functions but also richer forms of visualization, such as gallery-based, map-based, or timeline-based views, to further enhance the ways users can explore and engage with the educational materials.

Second, the methodology should be evaluated to determine how well it maximizes the utility of structured linked data while minimizing manual efforts to establish semantic links. This includes improving connectivity to specific digital archive content and introducing mechanisms such as automated link recommendations. We plan to evaluate the practical usefulness of the LOD infrastructure in real educational settings, including field trials with teachers and students using the developed tools and datasets. We also collaborate with projects such as S×UKILAM [41], which explore curriculum-resource linkages. In addition, a large-scale questionnaire survey on the perceived usefulness of the LOD infrastructure for archive-based learning is in preparation.

Third, it will also be important to assess the effectiveness of this framework in international contexts, particularly in linking educational metadata and content distributed outside of Japan. Although we have not tested other countries' curricula, our method — linking structured curriculum statements to external learning resources — is potentially generalizable to similar structured educational systems. With structured or semi-structured formats (e.g., standardized objectives), similar vocabularies could enable adaptation. However, manual alignment remains a common challenge. Shared infrastructures and reusable vocabularies will be essential for broader adoption.

Finally, clarifying the roles and responsibilities of stakeholders, such as educators, content developers, and policymakers, is essential. This involves building consensus on data quality and determining appropriate platforms for managing and maintaining structured educational data.

**Resource Availability Statement** The datasets we constructed are available from two LOD sites: Textbook LOD (https://w3id.org/jp-textbook/) and Course of Study LOD (https://w3id.org/jp-cos/). Teaching Unit LOD is provided as part of the Textbook LOD. The sources of datasets are also deposited on Zenodo for archival purposes at https://doi.org/10.5281/zenodo.15389358 and https://doi.org/10.5281/zenodo.15389408, respectively, for long-term accessibility.

**Acknowledgments.** This work was supported by JSPS KAKENHI Grants JP23K28384 and JP23K11761.

## References

1. Japan Search, https://jpsearch.go.jp/, last accessed 2025-04-03
2. Japanese Course of Study LOD, https://w3id.org/jp-cos/, last accessed 2025-04-08
3. Japanese Textbook LOD, https://w3id.org/jp-textbook/, last accessed 2025-04-08
4. NHK for School API, https://school-api-portal.nhk.or.jp/, last accessed 2025-04-20
5. Schema.org, https://schema.org/, last accessed 2025-04-17
6. Teaching With Europeana, https://teachwitheuropeana.eun.org/, last accessed 2025-04-11
7. Visualization of relationships among multiple textbooks, https://edudata-plus.github.io/visualize-textbook-unit/, last accessed 2025-04-18
8. Wikidata, https://wikidata.org/, last accessed 2025-07-31
9. Topics: linked data for libraries, archives, and museums. Information Standards Quarterly **24**(2/3), 2–50 (2012). https://doi.org/10.3789/isqv24n2-3.2012.01
10. Shapes Constraint Language (SHACL): W3C Recommendation 20 July 2017 (2017), https://www.w3.org/TR/shacl/, last accessed 2025-04-08

11. IEEE Standard 1484.12.1-2020: IEEE Standard for Learning Object Metadata. IEEE (2020). https://doi.org/10.1109/IEEESTD.2020.9262118
12. Teaching with Primary Sources: Looking at the Support Needs of Instructors. Ithaka S+R (2021), research Report. https://doi.org/10.18665/sr.314912
13. Abu-Salih, B., Alotaibi, S.: A systematic literature review of knowledge graph construction and application in education. Heliyon **10**(3), e25383 (2024). https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e25383
14. American Memory User Evaluation Team: American Memory User Evaluation: 1991-1993. Library of Congress (1993)
15. Ariyama, Y., Oi, M., Ako, T., Egusa, Y., Enomoto, S., Takaku, M.: Survey on ict use by teachers and staff to promote the use of digital archives in school education. Shiga Bunkyo Junior College Kiyo (27), 1–11 (2025). https://doi.org/10.32125/0002000056, (In Japanese)
16. Arms, C.R.: Historical collections for the national digital library: Lessons and challenges at the library of congress [part 1]. D-Lib Magazine **2**(4) (1996), https://www.dlib.org/dlib/april96/loc/04c-arms.html
17. Arms, C.R.: Historical collections for the national digital library: Lessons and challenges at the library of congress [part2]. D-Lib Magazine **2**(5) (1996), https://www.dlib.org/dlib/may96/loc/05c-arms.html
18. Baker, T., Bermès, E., Pompidou, C., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., Schneider, J., Singer, R., Summers, E., Waites, W., Young, J., Zeng, M.: Library Linked Data Incubator Group Final Report: W3C Incubator Group Report 25 October 2011 (2011), https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/
19. Barker, P., Campbell, L.: Metadata for learning materials: an overview of existing standards and current developments. Technology, Instruction, Cognition and Learning **7**(3-4), 225–243 (2010)
20. Bellin, A.: The fun of discovery: Library welcomes american memory fellows. Library of Congress Information Bulletin **58**(9), https://loc.gov/loc/lcib/9909/fellows.html
21. Brickley, D., Guha, R.: RDF Schema 1.1: W3C Recommendation 25 February 2014 (2014), https://www.w3.org/TR/rdf11-schema/, last accessed 2025-04-08
22. Davis, E., Heravi, B.: Linked data and cultural heritage: A systematic review of participation, collaboration, and motivation. Journal on Computing and Cultural Heritage **14**(2) (2021). https://doi.org/10.1145/3429458, https://doi.org/10.1145/3429458, https://doi.org/10.1145/3429458
23. DCMI Usage Board: DCMI Metadata Terms (2020), https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, last accessed 2025-04-08
24. Digital Archives Japan Promotion Committee/Practitioner Review Committee: Towards the realization of the digital archive society that japan aims for: Three-year summary report (2020), https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/r0208_3kanen_houkoku_honbun.pdf, last accessed 2025-04-10
25. Duerst, M., Suignard, M.: RFC 3987: Internationalized Resource Identifiers (IRIs). RFC 3987, Internet Engineering Task Force (Jan 2005), https://www.rfc-editor.org/rfc/rfc3987
26. Education Data Plus Research group: (u) understand how japanese-style culture was born by learning from the lifestyles and culture of the aristocracy. Course of Study LOD, https://w3id.org/jp-cos/8220263213000000

27. Egusa, Y., Takaku, M.: Building and Publishing Japanese Textbook Linked Open Data. The Journal of Information Science and Technology Association **68**(7), 361–367 (2018). https://doi.org/10.18919/jkg.68.7_361, (in Japanese)
28. Egusa, Y., Takaku, M.: A Case Study of Japanese Textbook Linked Open Data: Publishing a Small Bibliographic Collection from a Special Library. In: Proceedings of International Conference on Dublin Core and Metadata Applications 2019. pp. 75–79 (2018). https://doi.org/10.23106/dcmi.952141752
29. Gaitanou, P., Andreou, I., Sicilia, M.A., Garoufallou, E.: Linked data for libraries: Creating a global knowledge space, a systematic literature review. Journal of Information Science **50**(1), 204–244 (2024). https://doi.org/10.1177/01655515221084645, https://doi.org/10.1177/01655515221084645
30. GitHub: What is github pages?, https://docs.github.com/en/pages/getting-started-with-github-pages/what-is-github-pages, last accessed 2025-04-19
31. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org: evolution of structured data on the web. Communications of the ACM **59**(2), 44–51 (Jan 2016). https://doi.org/10.1145/2844544
32. Jones, E., Seikel, M.: Linked data for cultural heritage. An ALCTS monograph, ALA Editions (2016)
33. Komori, K.: A proposal for using digital archives in the school education. Journal of the Japan Society for Digital Archive **3**(2), 211–212 (2019). https://doi.org/10.24506/jsda.3.2_211, (In Japanese)
34. Maekawa, M.: A knowledge-recycling media environment through area-based digital commons: Support for area-based archiving through an open platform. Journal of the Japan Society for Digital Archive **7**(s2), s55–s58 (2023). https://doi.org/10.24506/jsda.7.s2_s55
35. Malta, M.C., Baptista, A.A., Walk, P.: Developing metadata application profiles. Advances in web technologies and engineering (AWTE) book series, Information Science Reference (2017). https://doi.org/10.4018/978-1-5225-2221-8
36. MEXT: Mext educational data standard, https://www.mext.go.jp/a_menu/other/data_00001.htm, last accessed 2025-04-08
37. MEXT: Textbook compiling prospectus, https://www.mext.go.jp/a_menu/shotou/kyoukasho/tenji/1364484.htm, last accessed 2025-04-08
38. Miles, A., Bechhofer, S.: Skos simple knowledge organization system reference (2009), https://www.w3.org/TR/skos-reference/
39. National Diet Library: NDL Search, https://ndlsearch.ndl.go.jp/en/, last accessed 2025-07-31
40. NHK: NHK for School, https://www.nhk.or.jp/school/, last accessed 2025-04-18
41. Oi, M.: S×UKILAM: Primary Source Sets / S×UKILAM Collaboration: A teaching materials archive using a variety of materials, https://adeac.jp/adeac-lab/top/SxUKILAM/, last accessed 2025-04-11
42. Oi, M.: Possibilities and challenges around the use of digital archives in education: practical examples. Current Awareness (352), 2–4 (2022), https://current.ndl.go.jp/ca2018
43. Oi, M., Kim, B., Watanave, H.: “S×UKILAM” collaboration to connect local digital resources and school education: Workshop and archiving to construct network of “people” and “data”. In: From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries. ICADL 2022. Lecture Notes in Computer Science, vol. 13636, pp. 125–134. Springer (2022), https://doi.org/10.1007/978-3-031-21756-2_10

44. Oi, M., Miyata, S., Ono, K., Ohmukai, I., Watanave, H.: Curation learning model utilizing digital archives: Connecting "questions" and "resources" in inquiry learning. Journal of the Japan Society for Digital Archive **7**(1), e1–e9 (2023). https://doi.org/10.24506/jsda.7.1_e1, (In Japanese)
45. Segawa, Y., Kigoshi, M., Takai, C., Takemura, H., Yokoyama, M., Takahashi, N.: Toward the educational utilization and apply of digital archive. Journal of College and University Libraries **121**, 2136 (2022). https://doi.org/10.20722/jcul.2136
46. Sutton, S., Golder, D.: Achievement Standards Network (ASN): An Application Profile for Mapping K-12 Educational Resources to Achievement Standards. In: Proceedings of International Conference on Dublin Core and Metadata Applications, DC-2008 (2008). https://doi.org/10.23106/dcmi.952109190
47. Takaku, M.: ttl2html, https://github.com/masao/ttl2html/, last accessed 2025-04-19
48. Tavakoli, M., Elias, M., Kismihók, G., Auer, S.: Metadata analysis of open educational resources. In: LAK21: 11th International Learning Analytics and Knowledge Conference. pp. 626–631. LAK21 (2021). https://doi.org/10.1145/3448139.3448208
49. Textbook LOD Project: Roadmap for Textbook LOD, https://github.com/jp-textbook/jp-textbook.github.io/wiki/Roadmap, last accessed 2025-04-18
50. Tokyo Gakugei University Library: Open Educational Contents Archive of Tokyo Gakugei University, https://d-archive.u-gakugei.ac.jp/en, last accessed 2025-05-13
51. Veccia, S.: American memory fellows institute ndl learning center hosts annual gathering. Library of Congress Information Bulletin **59**(10), https://loc.gov/loc/lcib/0010/amfellow.html
52. Vlachou, E., Deligiannis, I., Karydis, I.: Museum education using XR technologies: A survey of metadata models. European Journal of Engineering and Technology Research **1**(CIE), 66–77 (2023). https://doi.org/10.24018/ejeng.2023.1.CIE.3139
53. W3C Permanent Identifier Community Group: Permanent identifiers for the web, https://w3id.org/, last accessed 2025-04-18