

A Case Study of Japanese Textbook Linked Open Data: Publishing a Small Bibliographic Collection from a Special Library

Yuka Egusa
National Institute for
Educational Policy
Research, Japan
yuka@nier.go.jp

Masao Takaku
University of Tsukuba,
Japan
masao@slis.tsukuba.ac.jp

Abstract

Japanese Textbook Linked Open Data (LOD) is an LOD dataset of bibliographic and educational information that has been organized over the years by the Library of Education at the National Institute for Educational Policy Research. The dataset consists of bibliographic information for 7,548 volumes of Japanese textbooks authorized from 1992 to 2017, and provides 219,018 Resource Description Framework (RDF) triples as of April 2019. This paper reports a case study of the development and publication of Japanese Textbook LOD.

Keywords: textbook, Linked Open Data, bibliographic information, Uniform Resource Identifier

1. Introduction

Linked Open Data (LOD) has been gaining interest as a way to share structured data on the Web. LOD has been proposed as a framework for the semantic web as it allows us to deal with data from diverse disciplines with a variety of granularity (Heath & Bizer, 2011). In the LOD era, metadata that was traditionally organized and managed by libraries is of great interest. Many GLAM (galleries, libraries, archives, museums) institutions around the world have been trying to build and publish their collections as LOD, including not just bibliographic information for books and articles, but also authority data such as subject headings and name authorities (Baker et al., 2011).

The authors have built a structured dataset of bibliographic information of textbooks and published it as “Japanese Textbook Linked Open Data (LOD)” since January 2017 at <https://w3id.org/jp-textbook/>. Hereafter, we call this dataset “Textbook LOD.” The dataset was compiled from bibliographic data organized over an extended period of time by a special library, the Library of Education at National Institution for Educational Policy Research in Tokyo, Japan. The dataset consists of bibliographic information for 7,548 Japanese textbook volumes from 1992 to 2017, and provides 219,018 Resource Description Framework (RDF) triples as of April 2019.

Textbook LOD allows users to uniquely identify a textbook via an URI (Uniform Resource Identifier) and to obtain information about textbooks and related entities. Through Textbook LOD, textbooks are easily identifiable by using four elements: type of school, textbook symbol, textbook number, and year of authorization. It becomes easy to identify a textbook since these elements are always printed on the cover or colophon of the textbook. LOD technologies enable the retrieval of various details, not only bibliographic information like the title and publisher of the textbook, but also curriculum guidelines, subject area, subject, and so on.

2. Textbooks in Japan

In Japan, students in elementary, and lower secondary and upper secondary schools, from the age of 6 to 18, study the same standard syllabus based on curriculum guidelines (The Central Council for Education, 2008) published by the Ministry of Education and Science Technology and Sport. Textbooks for each subject provide specialized learning content in daily school classes.

A considerable amount of effort and resources have been invested in producing and authorizing these textbooks. In the textbook authorization system in Japan, the entire process, including planning and editing, screening by experts, authorization by the government, and selection by schools, usually takes approximately five or more years. The Japanese government provides free authorized textbooks to all elementary and junior high school students. Textbooks form the basis of daily learning and facilitate quality assurance and standardization of the national educational system (OECD, n.d.). The total number of unique textbook titles in the academic year 2018 was 897. The total number of volumes of textbooks in Japan for 2018 was approximately 124 million.

Textbooks also play a role by functioning as a generational memory, via which an entire generation shares common educational materials and content during their school lives. Patrons' demand for textbooks in libraries is also shown in reference and other services. For instance, the demand for textbooks by many users of reference services at the Library of Education was recorded and accounted for 43% (278 records) of the total records from 2008 to March 2018.

3. Basic URI Design

Textbook LOD assigns URIs to textbook resources and other related resources, such as curriculum guidelines, subject areas, subject, and textbook catalogues. These resources are handled as linked data resources. In RDF, we use an internationalized version of URI, "IRI" (Internationalized Resource Identifier). It is especially important for URIs to be internationalized because URIs in Textbook LOD contain Japanese character strings.

The base URI for Textbook LOD is "https://w3id.org/jp-textbook/," and hereafter we use the prefix "textbook:" to this base URI. This base URI is hosted at the permanent URI service operated by the World Wide Web Consortium Permanent Identifier Community Group (W3C Permanent Identifier Community Group, n.d.). As of April 2019, the actual website for Textbook LOD is hosted at Github pages and is available under <https://jp-textbook.github.io/>. Access to permanent URIs, starting with "https://w3id.org/jp-textbook/," is automatically redirected to the actual resources at the Github pages and provides useful dataset information, as described below.

4. Metadata model for Textbook LOD

Figure 1 shows an overview of metadata model for major resources with an example of a textbook in Textbook LOD.

The representation "textbook:中学校/2001/英語/904" depicted at the center of Figure 1 represents a textbook resource. As shown in Figure 1, the textbook resource has relationships with bibliographic literal values, such as the title literal value "NEW CROWN ENGLISH SERIES 3." In addition, it has relationships with specific textbook information such as the literal value "2001" as the authorized year and the literal value "2002-2005" as the usage years. There are also other associated resources, such as the subject area of the textbook "textbook:curriculum/中学校/2002/外国語, the subject "textbook:curriculum/中学校/2002/外国語/英語," and the curriculum guideline corresponding to the textbook "textbook:curriculum/中学校/2002." There are other relationships between the subject area and the curriculum guideline resources other than the textbook.

For the property vocabulary of Textbook LOD, we decided to use vocabulary that was as widely used as possible. For bibliographic information, most parts of the model use Schema.org vocabulary (schema.org, n.d.), which is accepted by major search engine companies and is actively maintained on the Web with a broad scope for coverage as LOD vocabulary. As properties of Schema.org, elements such as the book title (schema:name), editor's name (schema:editor), and publisher (schema:publisher) were used. On the other hand, as bibliographic properties are not found in Schema.org properties, the total number of pages (bf:extent) and sizes (bf:dimensions) was described using BIBFRAME vocabulary (BIBFRAME 2.0 Vocabulary,

2016). Other textbook-specific properties were described in our own namespace, including the authorized year (textbook:authorizedYear), usage years (textbook:usageYear), grade, subject area (textbook: subjectArea), subject (textbook:subject), and so on. In addition, the property representing the holding information of the Library of Education also has its own vocabulary in the namespace “http://dl.nier.go.jp/library/vocab/.”

Furthermore, the model’s constraints are described and documented via a shape expression language, Shapes Constraint Language (SHACL) (Knublauch & Kontokostas, 2017).

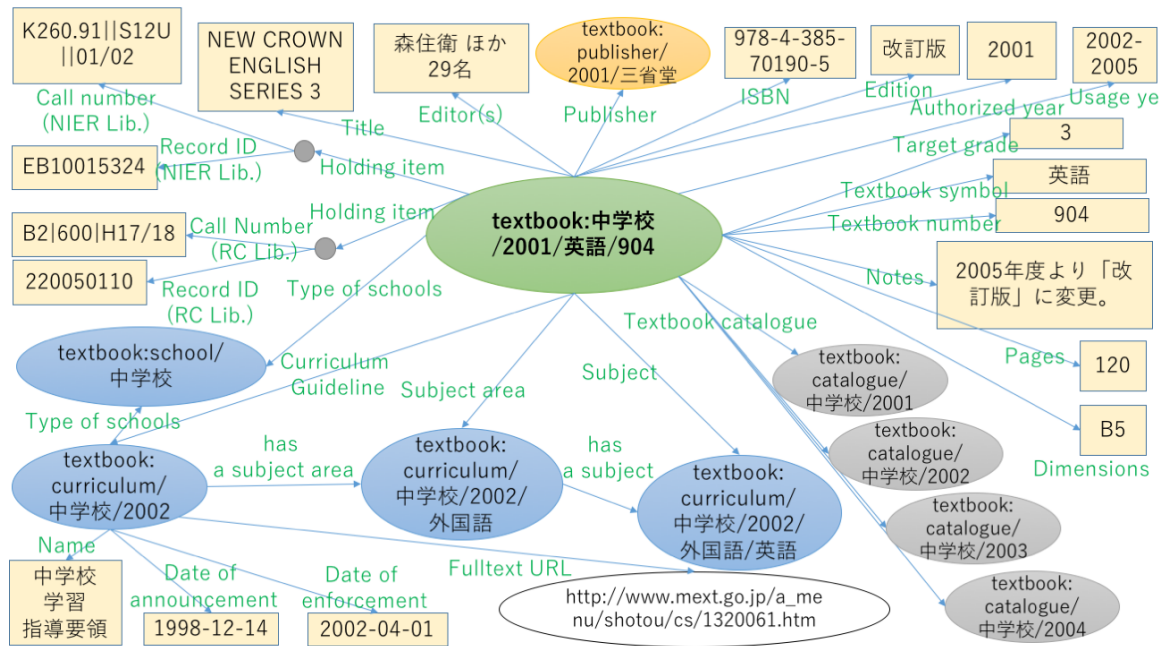


FIG. 1. Overview of metadata model for Textbook LOD

5. Hosting Textbook LOD

Textbook LOD is not just a dataset, but also a website that provides useful information through its URIs, according to the LOD principle outlined by Berners-Lee (2006).

When a user wants to access information about a textbook, he or she can obtain basic information through the corresponding URI of the textbook. For example, Figure 2 (a) shows the screenshot from the URI “https://w3id.org/jp-textbook/中学校/2011/公民/921.” Note that when we use the permanent URI service, the actual URI accessed is “https://jp-textbook.github.io/中学校/2011/公民/921.” As shown in the URI, this is a textbook on the subject “Civics Area” (公民) in junior high school (中学校) authorized in 2011. Furthermore, looking at the page, we can get textbook-specific information such as the book title, “New society citizen”; the editors’ names, “Fumihiko Gomi, Koji Toba, Noritaka Yagasaki, and 46 others”; the publisher’s name, “Tokyo Shoseki”; ISBN (international standard book number), “978-4-487-12048-2”; subject area, “Social Studies”; subject, “Social Studies (Civics Area)”; grade used, “third grade in junior high school”; years the textbook was used, “from 2012 to 2015”; and corresponding curriculum guideline, “2012 edition.” In addition, there are related links to the bibliographic details page of the relevant textbook in the Library of Education OPAC (Online Public Access Catalogue), results of National Digital Library (NDL) search, and CiNii Books search through the textbook’s ISBN.

(a) Example of a textbook page

```

@prefix textbook: <https://w3id.org/jp-
textbook/>.
@prefix bf:
<http://id.loc.gov/ontologies/bibframe/>.
@prefix schema: <http://schema.org/>.
@prefix rdfs:
<http://www.w3.org/2000/01/rdf-schema#>.

```

```

textbook:中学校/2011/公民/921
  a textbook:Textbook;
  bf:dimensions "AB";
  bf:extent "222";
  schema:editor "五味文彦 戸波江二 矢ヶ
崎典隆 ほか 46名";
  schema:isbn "978-4-487-12048-2";
  schema:name "新しい社会 公民";
  schema:publisher
    textbook:publisher/2011/東書;
  rdfs:seeAlso
    <http://id.ndl.go.jp/bib/023680706>,
    <http://id.ndl.go.jp/jpno/22098211>,
    <https://ci.nii.ac.jp/ncid/BB08897772>;
  textbook:authorizedYear "2011";
  textbook:curriculum textbook:curriculum/
  中学校/2012;
  textbook:grade 3;

```

(b) Example of an RDF/Turtle format

FIG. 2. Examples of textbook information

Information for a specific textbook can be obtained not just through a human-friendly web page but also in a machine-readable format, conforming to the LOD principle (Berners-Lee, 2006). When a content-negotiated request is sent to a URI, the service returns an RDF/Turtle format (Beckett, Berners-Lee, Prud'hommeaux, & Carothers, 2014), as shown in Figure 2 (b).

On the top page of Textbook LOD at “<https://w3id.org/jp-textbook/>” shown in Figure 3, there are several useful links to lists of subject areas and subjects that allow users to browse textbook information. There are also download links of the datasets as a format of RDF/Turtle.

In addition, there is detailed documentation about the dataset including its metadata model and revision history (Japanese Textbook LOD Project, n.d.).

6. Conclusion and Future Works

Textbook LOD is not very large, as it contains only about 7,000 volumes. The bibliographic information of the materials originally organized by the library has been reorganized and structured so that the resources can be shared as an LOD. Although the LOD dataset is hosted and maintained by the authors as a research project, we have continuously updated the dataset over two years in cooperation with the Library of Education. And one of the benefits of the LOD publishing is continuous data cleaning all over the dataset. The dataset is freely available under the CC0 (Creative Commons Zero) license.

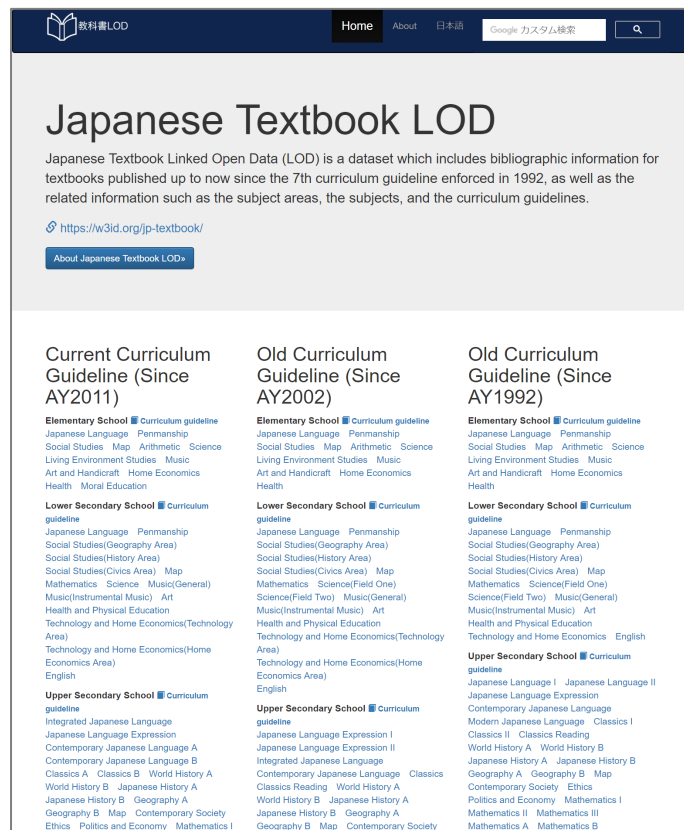


FIG. 3. Screenshot from the top page of Textbook LOD

There are several issues remaining for future work, including expansion of textbook coverage, more granular-level organization such as TOC (table of contents) and full-text content levels, and linking to additional external textbook identifiers.

References

- Baker, T., Bermes, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., . . . Zeng, M. (2011). Library Linked Data Incubator Group Final Report. Retrieved from <http://www.w3.org/2005/Incubator/11d/XGR-11d-20111025/>
- Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. Retrieved from <https://www.w3.org/TR/turtle/>
- Berners-Lee, T. (2006). Linked data. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>
- BIBFRAME 2.0 Vocabulary. (2016). Retrieved from <http://id.loc.gov/ontologies/bibframe/>
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space (No. 1). Morgan & Claypool.
- Japanese Textbook LOD Project. (n.d.). About Japanese Textbook LOD. Retrieved from <https://w3id.org/jp-textbook/en/about>
- Knublauch, H., & Kontokostas, D. (Eds.). (2017). Shapes Constraint Language (SHACL). Retrieved from <https://www.w3.org/TR/shacl/> (W3C Recommendation 20 July 2017)
- OECD. (n.d.). the OECD's Programme for International Student Assessment (PISA). Retrieved from <http://www.oecd.org/pisa/>
- schema.org. (n.d.). Retrieved from <http://schema.org>
- The Central Council for Education. (2008). The Improvements to the Courses of Study for Elementary, Lower Secondary and Upper Secondary Schools. In (chap. 1. Educational Objectives and Previous Revisions to the Courses of Study). Retrieved from http://www.nier.go.jp/04_kenkyu_annai/pdf/improvement_of_the_courses_of_study.pdf
- W3C Permanent Identifier Community Group. (n.d.). Permanent Identifiers for the Web. Retrieved from <https://w3id.org>