Improvement of metadata interoperability for promoting distribution and utilization of Japanese digital cultural resources

Daichi Machiya
National Diet Library,
Japan
d-machiy@ndl.go.jp

Michiko Takahashi
National Diet Library,
Japan
mmmichi@ndl.go.jp

Yuka Koguchi
National Diet Library,
Japan
koguti@ndl.go.jp

Abstract

The National Diet Library, Japan (NDL) has contributed to promoting utilization of Japanese digital cultural resources by operating Japan Search, the national metadata platform of Japanese digital content in various fields. While Japan Search accepts metadata regardless of schema to minimize the workload of data providers and enable rapid expansion of searchable databases, it is necessary to improve metadata interoperability to enable efficient utilization of resources.

This presentation covers practices to improve metadata interoperability such as an initiative in the community providing metadata to Japan Search and improvement of RDF provided by Japan Search.

Keywords: National Diet Library; Linked Open Data; metadata aggregation; metadata model; digital cultural heritage

1. Japan Search and Metadata Distribution

Japan Search is a national platform for aggregating metadata of Japanese digital resources of various fields officially launched in August 2020. It is operated by the NDL, in cooperation with a variety of organizations in Japan, under the policies established by the Working Group Steering Committee of the Digital Archives Japan Promotion Committee.

As of April 2022, Japan Search aggregates metadata of more than 25 million items from 171 databases and provides an integrated search service and the infrastructure for utilization of these databases.

Japan Search emphasizes utilization of resources and provides several functions, such as a "My Gallery" function which enables users to publish their own curated selection of resources, a search filter for use conditions of content, and an integrated IIIF viewer.

Japan Search collects metadata from aggregators for each field/region, and the NDL plays the role of the aggregator for the library field by providing the discovery service of NDL Search as well as acting as the operator of Japan Search.

Japan Search supports uploading metadata in a general file format such as Excel or TSV by data providers, regardless of metadata schema, in addition to OAI-PMH API. It gives the benefit of minimizing the partner's workload and enabling rapid expansion of searchable databases.

On the other hand, there is a need for interoperability to utilize digital cultural resources. Therefore, the NDL and the libraries of universities and research institutions have cooperated to develop guidelines to improve metadata interoperability. In addition, Japan Search provides metadata as RDF, and the NDL has continuously improved it to enable users to effectively retrieve information from Japan Search.

2. The Guidelines for Metadata Distribution in Japan

The NDL provides NDL Search as a discovery service for the holdings and digital collection of libraries and academic institutions in Japan. The NDL has steadily established a system to collect

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications

metadata of digitized library holdings and information resources. As the aggregator of the library field, the NDL provides aggregated metadata to Japan Search via the API of NDL Search.

NDL Search and other NDL web services use DC-NDL, a metadata schema based on Dublin Core. DC-NDL has also been used as the standard schema for metadata exchanges of union catalogs and digital collections of public libraries in Japan.

One of the challenges is aggregating metadata from digital archives of universities and research institutions, or databases providing digitized books and various cultural and research resources. The NDL and stakeholders have discussed this issue, and developed the Guidelines for Metadata Distribution in Japan.

In Japan, while metadata distribution of research results such as journal articles has been established, metadata distribution for academic digital archives has yet to be established.

Since some databases are developed by libraries together with researchers and projects outside the library's institution, academic libraries are expected to commit to distributing metadata of these databases using their experience.

Because a number of academic digital collections are published on institutional repositories (IRs), and delivering metadata from IRs to NDL Search is already implemented, it is expected to use this flow for efficient metadata distribution to Japan Search.

However, the metadata schema for IRs is insufficient to describe digitized contents because the main target of the schema had been journal articles. It has been recognized that there is a necessity to update the schema.

IRs delivers metadata in JPCOAR Schema, published as the metadata schema for research results in 2017. It was developed by the Japan Consortium for Open Access Repositories, consisting of Japanese academic libraries, and uses the vocabularies of Dublin Core, DataCite, and OpenAIRE. It focuses on international interoperability and sustainability.

The NDL and academic libraries have discussed the interoperability of two metadata schema, DC-NDL and JPCOAR Schema.

Based on discussions between the NDL and academic libraries, a public draft of the guidelines was published in March 2022. These guidelines show suitable metadata distribution flows, necessary common metadata items, mapping among schemas, and future update plans of each metadata schema. These guidelines also encourage academic libraries to provide metadata of their databases to adapt to current needs for discoverability and interoperability.

With further discussion and feedback on these guidelines, an update of DC-NDL and JPCOAR Schema to improve description of digital archives is planned. We plan to start providing metadata of digital collections on IRs to Japan Search via NDL Search and keep on communicating to optimize metadata distribution and expand it to other library communities.

3. Improvement of Japan Search RDF

Japan Search provides both metadata in the schema which the data provider delivered and as Japan Search RDF (JPS-RDF), metadata converted to the common RDF schema designed for universal utilization via SPARQL Endpoint.

JPS-RDF is designed for interoperability with external cultural heritage portals such as Europeana, so that it is possible to carry out global searches including them.

One of the features of JPS-RDF is the normalization of entities in metadata. Although information such as the name of a person (agents), time, place, subject and keyword is important for searching, in many cases, this information is described in a literal format in source databases. This causes insufficient search results in the case of spelling variants and namesakes. Normalization of entities as URI enables appropriate searches, and linking normalization URI to LOD hub enables enriched retrieves with outer identifiers and related information.

DC PAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications

JPS-RDF has continuously improved based on newly added databases and feedback from users. For instance, while the smallest unit of normalization for place was prefecture, now it is city or county, and it is linked to URI of the Statistical LOD of Japan, published by the Ministry of Internal Affairs and Communications.

Japan Search itself utilizes JPS-RDF for enhancing search functions, for example, providing detailed facets for search results, enriching item details with automatically displayed related items such as items created by the same creator and items created in same period.

Also, efforts are made to expand the utilization of Japan Search by importing external RDF which do not provide a SPARQL Endpoint to Japan Search's triple store. The representative example is NDC-LD, the linked data of Nippon Decimal Classification (NDC), one of the main library classification systems used in Japan. Some records in Japan Search include NDC classification, so linking each item to NDC-LD enables searching with conditions considering multiple versions of NDC and hierarchy of class.

Another effort for improvement is through contributing to Wikidata. Wikidata has a property for indicating the URI of a normalized entity in Japan Search (P6698), and some URIs are added by third parties. The NDL started to add unregistered URIs itself. This is expected to expand discoverability of digital resources on Japan Search and of related persons.

4. Conclusion

For the sake of optimizing before beginning to collect metadata, community discussions were held to create guidelines for interoperability. They shared suitable metadata flows and necessary common metadata items among libraries. The guidelines also show mapping between DC-NDL and JPCOAR Schema, and future update plans of each metadata schema.

RDF provided by Japan Search has been continuously improved by the normalization of entities and linking to outer LOD.

Efforts are planned to improve metadata interoperability, as well as to expand discoverable databases, for the sake of expanding the utilization of digital content with Japan Search.

Acknowledgements

References

Daichi Machiya, Tomoko Okuda, Masahide Kanzaki (2019). Japan Search RDF Schema: A Dual-Layered Approach to Describe Items from Heterogeneous Data Sources. Proceedings of the International Conference on Dublin Core and Metadata Applications, 2019, 21-25.

Masahide Kanzaki. Japan Search RDF Model Primer. Retrieved May 30, 2022, from https://www.kanzaki.com/works/ld/jpsearch/primer/.

National Diet Library, Japan. An Overview of the JPS-RDF Schema: Data that is obtainable from JPS SPARQL endpoints. Retrieved May 30, 2022, from https://jpsearch.go.jp/static/developer/introduction/en.html.

National Diet Library, Japan. National Diet Library Dublin Core Metadata Description (DC-NDL). Retrieved May 30, 2022, from https://www.ndl.go.jp/en/dlib/standards/meta/index.html.

National Diet Library, Japan et al. The Guidelines for Metadata Distribution in Japan (Public Draft). Retrieved May 30, 2022, from https://iss.ndl.go.jp/information/guideline/.

OOZONO Hayahiko, KATAOKA Tomoko, TAKAHASHI Nanako, TAGUCHI Tadasuke, HAYASHI Yutaka, MINAMIYAMA Yasuyuki. (2018). JPCOAR Schema: A brand-new metadata schema for smooth international distribution of scholarly output in Japan. Journal of Information Processing and Management, 60(10), 719-729. https://doi.org/10.1241/johokanri.60.719