Presentation
Japanese Metadata Standards "National Diet Library Dublin Core Metadata Description (DC-NDL)": Describing Japanese Metadata and Connecting Pieces of Data

Saho Yasumatsu
National Diet Library,
Japan
s-yasuma@ndl.go.jp

Akiko Hashizume
National Diet Library,
Japan
hasizume@ndl.go.jp

Julie Fukuyama
National Diet Library,
Japan
ju-fukuy@ndl.go.jp

Keywords: The National Diet Library; National library; Japanese library; metadata; Linked Data

1. About the NDL

The National Diet Library (NDL) is the sole national library in Japan. The NDL acquires, preserves and provides Japanese publications which are the nation's cultural and intellectual assets. The acquisition of library materials is mostly based on the Legal Deposit System. The NDL compiles and provides various bibliographies of library materials. Most of the collections are searchable through the NDL-OPAC and NDL Search on the website. To facilitate effective data use by computer systems or applications, the NDL initiatives to promote Linked Data and provides metadata as Linked Data.

This presentation shows the National Diet Library Dublin Core Metadata Description (DC-NDL), which defines elements and rules for the NDL's metadata. The views and opinions expressed in this presentation are those of the authors and do not necessarily represent the views or policies of the NDL or related organizations.

2. What is DC-NDL?

The DC-NDL is a descriptive metadata standard that is utilized primarily for converting catalog records of publications held by the NDL into metadata based on the Dublin Core Metadata Element Set (DCMES) and the DCMI Metadata Terms. The DC-NDL comprises three parts: NDL Metadata Terms, a list of metadata terms defined by the NDL; Application Profiles, which clarify the standard use of the DC-NDL in the NDL systems; and RDF Schema, which is a Resource Description Format (RDF) version of the NDL Metadata Terms.

In this presentation, the authors focus mainly on the Application Profiles. The DC-NDL is structured as shown in Fig. 1. It is a triple-layer data model, comprising Administrative Information, Bibliographic Information, and Item Information. This presentation contains an explanation of Bibliographic Information.

◉DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2016

```
<rdf:RDF>
  <dcndl:BibAdminResource rdf:about="http://iss.ndl.go.jp/books/R100000002-
I000002158611-00">
    <dcndl:catalogingStatus>C7</dcndl:catalogingStatus>
    <dcndl:record rdf:
I000002158611-00#materi
  </dcndl:BibAdminResou
```

Information about metadata management (e.g., cataloging status) → **Administrative Information**

```
  <dcndl:BibResource rdf:about="http://iss.ndl.go.jp/books/R100000002-
I000002158611-00#material">
    <dcterms:identifier rdf:datatype="http://ndl.go.jp/dcndl/
201869-2</dcterms:identifier>
    <dc:title>
      <rdf:Description>
        <rdf:value>謎の七支刀</rdf:valu
        <dcndl:transcription>ナゾ ノ シチシトウ</dcndl:transcription>
      </rdf:Description>
    </dc:title>
    <dcterms:creator>
      <foaf:Agent rdf:about="http://id.ndl.go.jp/auth/entity/00045529">
        <foaf:name>宮崎, 市定, 1901-1995</foaf:name>
        <dcndl:transcription>ミヤザキ, イチサダ</dcndl:transcription>
      </foaf:Agent>
    </dcterms:creator>
    <dcterms:publisher>
      <foaf:Agent>
        <foaf:name>中央公論社</foaf:name>
        <dcndl:location>東京</dcndl:location>
      </foaf:Agent>
    </dcterms:publisher>
    <dcndl:digitizedPublisher>国立国会図書館</dcndl:digitizedPublisher>
    <dcndl:dateDigitized>2013</dcndl:dateDigitized>
    <dcterms:subject>
      <rdf:Description rdf:about="http://id.ndl.go.jp/auth/ndlna/00568294">
        <rdf:value>日本--歴史--古代</rdf:value>
      </rdf:Description>
    </dcterms:subject>
    <dcterms:subject rdf:resource="http://id.ndl.go.jp/class/ndlc/GB38"/>
    <dcterms:subject rdf:resource="http://id.ndl.go.jp/class/ndc9/210.02"/>
  </dcndl:BibResource>
```

Information about the description of resources → **Bibliographic Information**

```
  <dcndl:Item rdf:about="http://iss.ndl.go.jp/books/R100000002
00#item">
    <dcndl:holdingAgent>
      <foaf:Agent>
        <foaf:name>国立国会図書館<
        <dcterms:identifier
rdf:datatype="http://ndl.go.jp/d                         s:identifier>
      </foaf:Agent>
    </dcndl:holdingAgent>
    <dcndl:callNumber>GB38-E5</dcndl:callNumber>
  </dcndl:Item>
</rdf:RDF>
```

Information about the description of holdings at each institution (e.g., call numbers, missing numbers) → **Item Information**

FIG. 1. Basic structure of the DC-NDL
*FIG. 1. does not show the full elements, but basic elements of the DC-NDL.

2.1. A Brief History

The DC-NDL was developed in order to enhance the interoperability of metadata among Japanese libraries and related institutions. It has been revised three times (in May 2007, June 2010 and December 2011) since its launch in March 2001, to add metadata terms and redefine their usages, reflecting revisions to the Dublin Core and the expansion of NDL services.

3. Features of the DC-NDL

The DC-NDL requires mechanisms that can support the NDL's own mission, such as collecting and preserving Japanese publications. In other words, the DC-NDL has to accommodate not only certain unique features of the Japanese language but NDL activities. At the same time, the DC-NDL needs to maintain the broad perspective of linking to the world's data.

The key functions of the DC-NDL are the follows: (1) Representing the *yomi* (pronunciation) of the Japanese language, (2) Connectivity with Linked Data, and (3) Compatibility with digitized materials.

3.1. Representing *yomi* (pronunciation): characteristics of the Japanese language

The Japanese language has three distinct types of characters: *hiragana* (cursive syllabary), *katakana* (angular syllabary) and *kanji* (Chinese character).

いちご ／ イチゴ ／ 苺

FIG. 2. Example of *hiragana, katakana,* and *kanji* for "strawberry."
*All of these are pronounced "i-chi-go"

Additionally, there are cases in Japanese where pronunciation will vary depending upon meaning, so it is necessary to indicate the pronunciation using *hiragana* or *katakana*.

″西瓜″
●Pronunciation 1: su-i-ka
●Pronunciation 2: ni-shi-u-ri

FIG. 3. Example of a word pronounced differently from its character representation
*Watermelon in Japanese could be pronounced either "su-i-ka" or "ni-shi-u-ri"

Furthermore, the Japanese language generally does not have spaces inserted between words. This makes it very difficult to parse correctly when the entire sentence is written phonetically in *hiragana* or *katakana*.

ココデハキモノヲヌイデクダサイ
●Pronunciation 1: ココ デ ハ キモノ ヲ ヌイデ クダサイ
        "Koko-de-wa-kimono (means clothes)-wo-nuide-kudasai"
●Pronunciation 2: ココ デ ハキモノ ヲ ヌイデ クダサイ
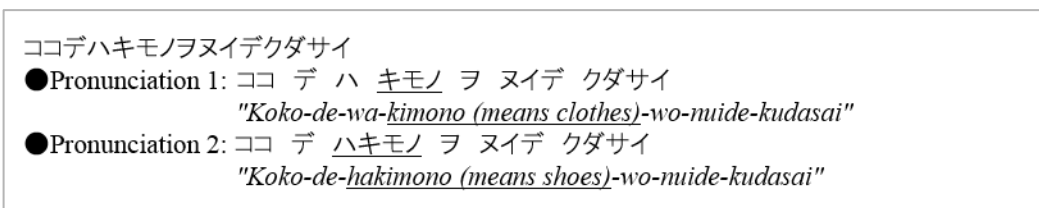   *"Koko-de-hakimono (means shoes)-wo-nuide-kudasai"*

FIG. 4. Example of a sentence with two different meaning, "clothes" or "shoes," depending on how the sentence is parsed.
*Pronunciation 1 means "Please take off your clothes here," Pronunciation 2 means "Please take off your shoes here."

The DC-NDL has been designed to serve as a national standard, and therefore defines metadata terms which can represent these characteristics of the Japanese language. Taking the Title field as an example, there are three salient features, as described below.

First, to describe the *yomi* (pronunciation) or transliteration of a particular title, the DC-NDL defines original properties such as a Transcription [dcndl:transcription], a Title Transcription [dcndl:titleTranscription], and an Alternative Transcription [dcndl: alternativeTranscription].

Second, the value stored in the Transcription [dcndl:transcription] property is the sentence with blanks between words, which is called *wakachi-gaki* in Japanese.

DC PAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2016

Third, the DC-NDL title representation and its pronunciation can be described as a set. Rather than using dcterms:title with restrictions on the literal range, dc:title is used without restrictions in its range. Describing a representation in a set with its *yomi* allows metadata to show semantic structures or logical relations of original catalog records, so that computers can handle the semantic structures and process them.

```
[Extract data from Bibliographic Information]
<dc:title>
  <rdf:Description>
    <rdf:value>謎の七支刀</rdf:value>
    <dcndl:transcription>ナゾ ノ シチシトウ</dcndl:transcription>
  </rdf:Description>
</dc:title>
```

FIG. 5. Example of a set comprising the representation and pronunciation of a title.

3.2. Connectivity with Linked Data

In the DC-NDL, Uniform Resource Identifier (URI) values can be used to link with other data. The following is an explanation of how to link an element to other data using a URI to link to creator and subject.

The DC-NDL creator property is intended to be used with URIs from Web NDL Authorities. Web NDL Authorities are a web service created by the authority data and maintained by the NDL. Web NDL Authorities include links to the Virtual International Authority File (VIAF), so that the authority data of the NDL can link to other major name authority files around the world through the VIAF.

```
[Extract data from Bibliographic Information]
<rdf:RDF>
  <dcndl:BibAdminResource
rdf:about="http://iss.ndl.go.jp/books/R100000002-I000002158611-00">
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    </dc:title>
    <dcterms:creator>
      <foaf:Agent rdf:about="http://id.ndl.go.jp/auth/entity/00045529">
        <foaf:name>宮崎, 市定, 1901-1995</foaf:name>
        <dcndl:transcription>ミヤザキ, イチサダ</dcndl:transcription>
      </foaf:Agent>
    </dcterms:creator>
    <dcterms:publisher>
      <foaf:Agent>
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  </dcndl:Item>
</rdf:RDF>
```

Web NDL Authorities
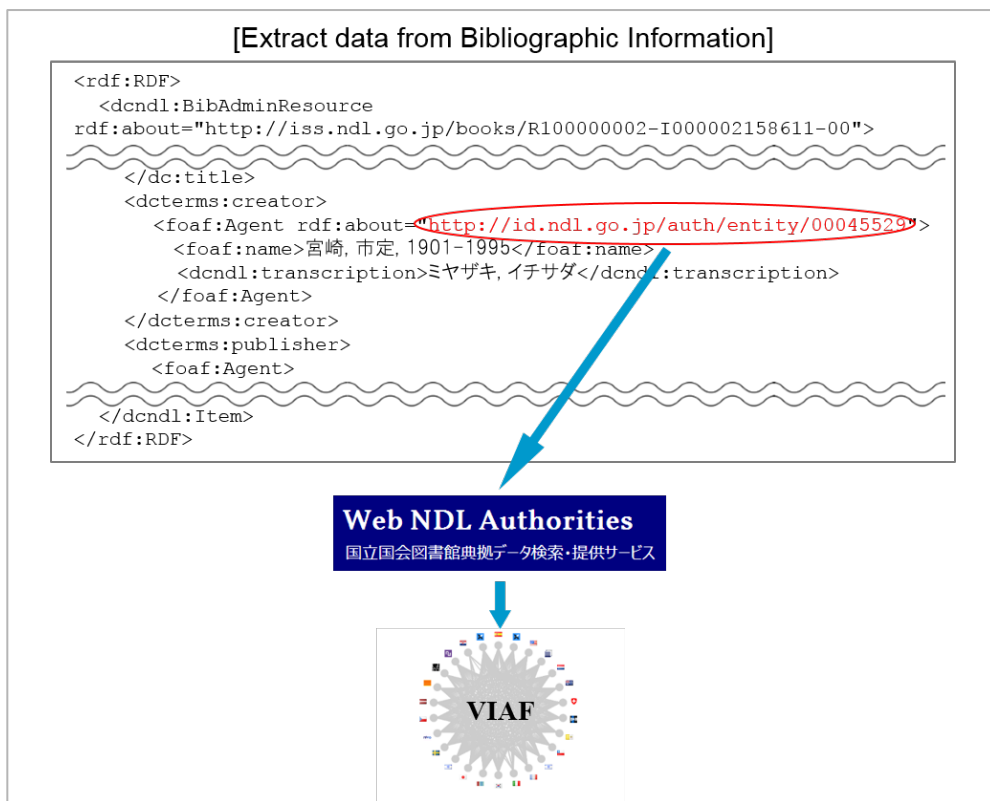国立国会図書館典拠データ検索・提供サービス

VIAF

FIG. 6. Example of link to "dcterms:creator"

The DC-NDL subject property is intended to use URI values from classification systems adopted by the NDL, such as the Nippon Decimal Classification (NDC) 9th edition, the National Diet Library Classification (NDLC) and the Dewey Decimal Classification (DDC). By the way, the NDC is a Japanese standard classification, and the NDLC is developed by the NDL.

```
[Extract data from Bibliographic Information]
●Example 1: The NDC 9th edition
  <dcterms:subject rdf:resource="http://id.ndl.go.jp/class/ndc9/210.02"/>

●Example 2: The NDLC
  <dcterms:subject rdf:resource="http://id.ndl.go.jp/class/ndlc/GB38"/>
```

FIG. 7. Example of using URIs in "dcterms:subject"
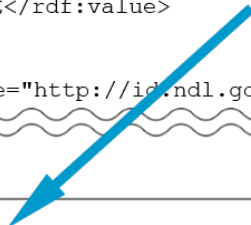
In addition, the subject property can use URIs from the National Diet Library List of Subject Headings (NDLSH). NDLSH, which is compiled and maintained by the NDL, is a controlled subject vocabulary for accessing information resources from a subject. Most NDLSH subject headings link to the Library of Congress Subject Headings (LCSH) as a part of the Web NDL Authorities. So that, via LCSH the NDLSH links to subject headings of several national libraries in the world.

```
[Extract data from Bibliographic Information]
<rdf:RDF>
  <dcndl:BibAdminResource rdf:about="http://iss.ndl.go.jp/books/R100000002-
I000002158611-00">
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    <dcndl:dateDigitized>2013</dcndl:dateDigitized>
    <dcterms:subject>
      <rdf:Description rdf:about="http://id.ndl.go.jp/auth/ndlna/00568294">
        <rdf:value>日本--歴史--古代</rdf:value>
      </rdf:Description>
    </dcterms:subject>
    <dcterms:subject rdf:resource="http://id.ndl.go.jp/class/ndlc/GB38"/>
  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  </dcndl:Item>
</rdf:RDF>
```

Web NDL Authorities
国立国会図書館典拠データ検索・提供サービス

LCSH
LIBRARY OF CONGRESS

FIG. 8. Example of link to "dcterms:subject"

3.3. Compatibility with digitized materials

The NDL in making every effort to digitize its holding. To help describe those digitized contents adequately, the DC-NDL also defines original metadata terms.
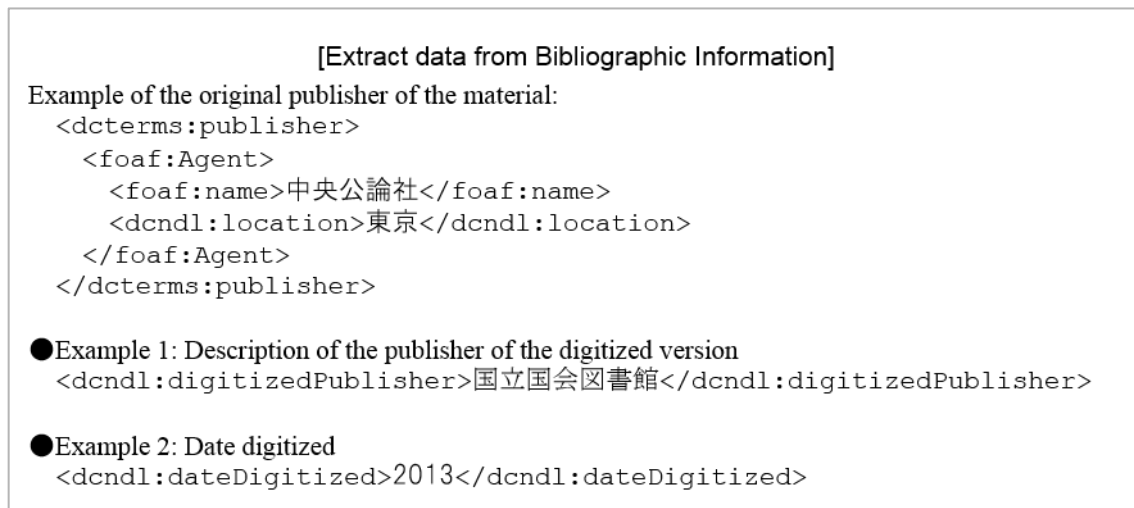
```
[Extract data from Bibliographic Information]
Example of the original publisher of the material:
  <dcterms:publisher>
    <foaf:Agent>
      <foaf:name>中央公論社</foaf:name>
      <dcndl:location>東京</dcndl:location>
    </foaf:Agent>
  </dcterms:publisher>

●Example 1: Description of the publisher of the digitized version
  <dcndl:digitizedPublisher>国立国会図書館</dcndl:digitizedPublisher>

●Example 2: Date digitized
  <dcndl:dateDigitized>2013</dcndl:dateDigitized>
```

FIG. 9. Examples of terms for digitized content

4. Using the DC-NDL: NDL Search

4.1. What is NDL Search?

Finally, here is an example of implementing DC-NDL for use with NDL Search.

NDL Search is an integrated information search service that serves as a gateway to the rich repository of knowledge contained in the NDL, public libraries, academic libraries, archives, museums, and academic research institutions in Japan. It officially opened to the public on January 2012 and can search about 85 million metadata records as of March 2016. Data sources for the NDL Search include: NDL-OPAC, Japanese Periodicals Index, National Diet Library Digital Collections, digital archives provided by public and academic libraries in Japan, etc.

4.2. The API and its terms of use

NDL Search provides bibliographic data in RDF/XML of books, journals, articles, newspapers, and digitized material, audio files, web pages, and other digital content. This data includes title, author, publisher, subject matter, classifications, ISBN, ISSN, National Bibliography No., NDLJP (the NDL Digital Collection), URLs of webpages which show digitized content (http://dl.ndl.go.jp/...), and information related to copyright protection. The NDL Search accesses metadata in two encoding formats, DC-NDL (RDF) and DC-NDL (Simple), which are available from the user interface as well as via the API. The API can be used with the following protocols; SRU, SRW, OpenSearch, Z39.50 and OAI-PMH. The amount of information each encoding format provides is varied.

As of July 15, 2016, free use of metadata is predicated on the assumption that the use is non-commercial. People who wish to use the API of the NDL Search for a commercial activity are requested to apply for a license. Also, people who require continuous access to the API for a non-profit activity are requested to apply for a license.

5. The DC-NDL in the future: for connecting more data

The following are issues for future research on the DC-NDL:

The DC-NDL creator and subject fields link to other data, such as LCSH. To make the DC-NDL more linkable, we need to implement a means to add links. Thus, we must examine which elements are appropriate for linking with other data and which URIs are suitable for including in our data. Moreover, we need to identify data that can be linked with ours.

References

DC-NDL. Retrieved July 15, 2016, http://www.ndl.go.jp/en/aboutus/standards/meta.html

Web NDL Authorities. Retrieved July 15, 2016, from http://id.ndl.go.jp/auth/ndla

VIAF. Retrieved July 15, 2016, from http://viaf.org/

NDL Search. Retrieved July 15, 2016, from http://iss.ndl.go.jp/?locale=en