JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

JLIS.it

"Yomi" (readings) in bibliographic data for materials in Japanese

Kazue Murakami[(a)]

a) National Diet Library, Japan

Contact: Kazue Murakami, kazue-m@ndl.go.jp
Received: 20 January 2022; **Accepted:** 21 March 2022; **First Published:** 15 May 2022

ABSTRACT
In bibliographic and authority data of materials in Japanese, it is necessary to represent kanji (Chinese characters) and its reading as a pair. Readings, called "yomi" in Japanese, are represented in katakana or the alphabet (romanization form). The readings in Japanese language have many variations depending on the context, so Japanese bibliographic information needs to provide both kanji and its reading as a pair to avoid misunderstandings.
The JAPAN/MARC is a machine readable version of the Japanese National Bibliography and provides sets of kanji, katakana-yomi and romaji-yomi (romanization form) of bibliographic and authority data. This paper introduces how readings of bibliographic and authority data in Japanese have been expressed and handled mainly in JAPAN/MARC, compared with international standards.

KEYWORDS
Japanese National Bibliography; JAPAN/MARC; National Diet Library Japan; Readings of Japanese; Romanization.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

Introduction

When creating bibliographic data for materials in Japanese, it is necessary to take into account the "yomi" (readings) specific to Japanese.
The Japanese language is expressed using kanji (Chinese characters), hiragana and katakana, which were uniquely developed in Japan based on kanji. There are usually at least two readings for each kanji in Japan: "on-yomi", which is the reading that came with the kanji from China, and "kun-yomi", which is a translation of the meaning of the kanji into Japanese. The reading of a kanji is usually expressed using hiragana or katakana.
In this paper, an explanation of yomi in Japanese in general will be given at first. And then, it will be introduced how yomi has been expressed mainly in the bibliographic data provision format of the National Diet Library, Japan (NDL).

1. What is "yomi" in Japanese?

In this section, an explanation about "yomi" in Japanese language in general is given with some examples before introducing expressions in bibliographic data.
The introduction mentioned that there are usually at least two readings for kanji: on-yomi and kun-yomi. For example, the kanji "大", which means "big", has two on-yomi: "dai" (e.g., "重大", pronounced as jūdai[1]) and "tai" (e.g., "大家", pronounced as taika), as well as kun-yomi: ō(kii), usually written with kanji and hiragana as "大きい".
Words with the same reading may also have different meanings depending on the kanji characters used. For instance, "大家" means a great expert, "対価" means payment, and "退化" means degeneration, but the readings for are all of these are "taika".
Furthermore, even words that combine the same kanji characters may have different meanings if they are read differently. For example, when the kanji "大家" is read as "taika", it means a great expert as mentioned above. However, when the same word is read as "ōya", it means a landlord.
Therefore, when Japanese words are expressed in kanji, it can often be misleading unless their readings are also presented.
As for personal names, various readings are possible even when the same kanji is used, and the same reading is not always used with the same kanji. For example, the family name "三田" is read as "Mita" or "Sanda" depending on the family. The family name "Itō" uses the kanji "伊東", "伊藤" or "井藤" depending on the family. There can be many variations of kanji and reading pairs for given names because more freedom is allowed for them than for family names.

[1] In this paper, pronunciations of Japanese words are expressed using the ALA Romanization tables. Japanese (2012), https://www.loc.gov/catdir/cpso/romanization/japanese.pdf.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

Table 1: One kanji with several pronunciations

Kanji	Yomi (Pronunciation/Romanization)
大	dai (だい)
	tai (たい)
	ō (おお)

Table 2: One kanji with several pronunciations (common nouns)

Kanji	Yomi (Pronunciation/Romanization)
大家	taika (たいか)
	ōya (おおや)

Table 3: One kanji with several pronunciations (proper nouns)

Kanji	Yomi (Pronunciation/Romanization)
三田	Mita (みた)
	Sanda (さんだ)

Table 4: One pronunciation with several kanji (common nouns)

Yomi	Kanji	Meaning
Taika	大家	Great expert
	対価	Payment
	退化	Degeneration

Table 5: One pronunciation with several kanji (proper nouns)

Yomi	Kanji
Itō	伊東
	伊藤
	井藤

In Japan, it is customary to include bibliographic information like a colophon at the end of a book. Since there are many variations of readings for the same kanji, the readings of the kanji for the title and the author are usually indicated to avoid misunderstandings in bibliographic information or an author's biography. These indications are useful to identify titles and creators, and to create authority data in cataloging processes.

In particular, in the process of creating authority data, even if the creator is indicated by kanji without its reading in the material, catalogers always assign its reading for the authorized access point by checking reference materials, inquiring with the publisher, or estimating the most common reading of the kanji.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

In the days when people used catalog cards, bibliographic data had to be listed in a certain order. In Japan, the most common way to list catalog cards was the order of readings of titles or authors. The pronunciation of hiragana and katakana is determined on a one-to-one basis for each character, with 48 basic characters each, while kanji is said to have tens of thousands of characters. It seemed that the most efficient way to arrange catalog cards was to use hiragana or katakana to represent the reading of the author's name or book title and list them in that order. That was why this way became the mainstream.

In catalog cards, personal name authority data was listed in the order of the readings of the personal name, while the kanji form was adopted as the heading, and identification was based on the kanji form and their readings. Therefore, it is customary not to consider it to be the same person if the kanji is different, even if the reading is the same. It is also customary not to regard the same kanji as the same name if it has a different reading, since different readings of the same kanji will be listed in different places.

Most libraries use online catalogs now, so it is no longer necessary to consider the sequence of catalog cards. However, as mentioned above, the same kanji can have different meanings when it is read in different ways, and even when read in the same way, the meanings can differ depending on the kanji used. Therefore, it is necessary to be able to search by both kanji and reading for the identification of Japanese materials.

Web NDL Authorities[2], which enables users to search for and acquire multiple authority data records, displays the kanji forms and katakana "yomi" forms of the authorized and variant access points, as well as the romanized "yomi" forms which replace the katakana "yomi" by alphabet. By providing both kanji and their readings, Web NDL Authorities can be used not only to identify individuals and works in combination with conventional kanji forms, but also as a basis for reading when creating reading materials for the visually impaired, for example.

2. Japanese National Bibliography and JAPAN/MARC

As the national bibliographic agency of Japan, the NDL has provided standard bibliographic data of domestic materials acquired by the legal deposit system and of domestic materials and Japanese materials published in foreign countries acquired by purchase and donation. The NDL also provides JAPAN/MARC as a machine-readable version of the Japanese national bibliography.

The printed "Japanese national bibliography weekly list" was a list of government publications sorted by which government agency published them, and private publications sorted by the Nippon Decimal Classification (NDC) system. The lists were in katakana-yomi order for each government agency and the NDC section, but the readings themselves were not provided. JAPAN/MARC provided katakana-yomi and romaji-yomi as well as kanji forms of bibliographies, so users became to be able to use readings for their searches and identify materials easier.

Before explaining how readings are provided in JAPAN/MARC, this section provides an overview of the history of JAPAN/MARC, and the next section will explain some Japanese romanization systems.

[2] https://id.ndl.go.jp/auth/ndla.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

At first, the NDL launched JAPAN/MARC(M) for monographs which provided bibliographies of books in 1981. It then started providing bibliographies of serials in 1988. In 2014, the two were merged into a single entity and renamed JAPAN/MARC (M/S). In 1997, the NDL also started providing JAPAN/MARC for authorities of authors' names. It was renamed JAPAN/MARC(A) in 2003. JAPAN/MARC for monographs, serials and authorities were all based on UNIMARC. Although there have been some changes such as addition of fields, the unification of formats for books (monographs) and serials in 2002, and the expansion of scope to include cartographic materials, audio-visual materials, etc. after 2003, the stance of complying with UNIMARC was not changed until the JAPAN/MARC MARC 21 format was implemented.
Since January 2012, both bibliographic data and authority data have been provided in the JAPAN/MARC MARC21 format. Currently, new, updated, and deleted data files are posted on the website[3] every week, and bibliographic data retrieved from the National Diet Library Search (NDL Search)[4] can be downloaded in MARC format.

3. Japanese romanization systems

Some examples will be given to illustrate the difference between the ALA-LC Romanization Table and the NDL for the Romanization of Japanese.
The ALA-LC Romanization Table uses macrons (¯) for the Japanese vowels a, i, u, e, and o to represent long vowels such as "ka a" in "o ka a sa n", which means mother. The NDL used circumflexes (^) for representing long vowels until March 2002. However, since April 2002, the NDL has not used any symbol such as a macron or circumflex. The example "o ka a sa n" is written as "okāsan" according to the ALA-LC Romanization Table, but in the NDL rules until March 2002, it was written "okâsan". In April 2002, it was decided to convert the katakana reading "**オカアサン**" into Roman characters one by one and write it as "okaasan". Since December 2011, the NDL has decided to represent this as "okasan" instead of converting katakana characters one by one. In the bibliographic data and authority data currently available by the NDL, all romaji-yomi are provided with the rules after December 2011.
In addition, there are two main types of Japanese Romanization: the Hepburn system, which emphasizes compliance with English pronunciation, and the Kunrei system, which is Japanese Cabinet-ordered. For example, "**ジリツ**," which means to control oneself, is written as "jiritsu" according to the ALA-LC Romanization Table, which is based on the Hepburn system. On the other hand, it is written as "ziritu" according to the Kunrei system. The NDL's bibliographic and authority data generate romaji-yomi from katakana-yomi systematically. The NDL began to record romaji-yomi based on the Hepburn system instead of the Kunrei system in December 2011.
For people who cannot read Japanese, romanization can be useful. However, as the NDL uses different rules for romanization depending on the time period, there are many different ways of romanizing Japanese. It is recommended to use several notations such as the Kunrei and Hepburn sytems when searching bibliographic or authority data by romaji reading.

[3] https://www.ndl.go.jp/jp/data/data_service/jnb_product.html.
[4] https://iss.ndl.go.jp/.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

4. Romanization in JAPAN/MARC

The JAPAN/MARC format has been designed to allow the expression of kanji form and reading form as a set, because, as mentioned in section 1, the ability to search by both kanji and reading forms is necessary for the identification of Japanese materials.
In this section, how the JAPAN/MARC format has recorded the kanji and reading forms in the fields will be explained, comparing them with the international standards to which the format adheres.

4.1 JAPAN/MARC (M) and (S) before 2011

In JAPAN/MARC (M), which was launched in 1981, the description blocks (2XX-3XX) were written in kanji characters, and the access point blocks (55X) contained a set of kanji and readings: katakana-yomi, and romaji-yomi forms. The format of JAPAN/MARC (M) was based on UNIMARC, however, its used some fields which were not specified in UNIMARC. For example, in UNIMARC (1977), the title and responsibility were recorded in field 200, but in JAPAN/MARC (M), the title and responsibility were recorded in field 251, which is not specified in UNIMARC. If there were more than one title, fields 252-259 were used. The edition statement and publication area were also recorded using fields not specified by UNIMARC. 5XX is the block for recording related titles, while JAPAN/MARC (M) used fields 551-559, which were also not specified by UNIMARC. $B for the kanji form, $A for katakana-yomi form, and $X for the romaji-yomi form were in the same field and in 2-byte characters, so that kanji and reading could be expressed as a set. Figure 1 shows that the title proper "函館・ロシアその交流の軌跡" was in 251$A, its katakana-yomi form "ハコダテ ロシア ソノ コウリュウ ノ キセキ" was in title heading (551$A) and its romaji-yomi "Hakodate rosia sono kouryuu no kiseki" was in 551$X. By recording "251A1" in 551B, it indicated that 551$A and $X were paired with 251$A.
This rule was also implemented in JAPAN/MARC (S) and JAPAN/MARC (M/S).

Title proper and statement of responsibility	251	$A	函館・ロシアその交流の軌跡
		$F	清水恵//著#
Parallel title proper	261	$A	Хакодатз Русские страницы в истории японского города#
Publication, distribution, etc.	270	$A	［函館］
		$B	函館日口交流史研究会
		$D	2005. 12#
Physical description	275	$A	387p
		$B	20cm#
Title heading	551	$A	ハコダテ ロシア ソノ コウリュウ ノ キセキ
		$X	Hakodate rosia sono kouryuu no kiseki
		$B	251A1
		$A	Khakodatz Russkie stranitsy v istorii iaponskogo goroda
		$X	Khakodatz Russkie stranitsy v istorii iaponskogo goroda
		$B	261A1#

Figure 1: A partial example of JAPAN/MARC (M/S) format (2009)

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

4.2 JAPAN/MARC (A) before 2011

In "JAPAN/MARC for authorities of authors' names", which was launched in 1996, the same field was repeated for each kanji, katakana-yomi, and romaji-yomi form. To indicate that it was a single set, a code indicating the relationship between the fields and a linking number to match the linked fields were recorded in $6. This method was based on that specified in UNIMARC/Authorities (1991). Figure 2 shows field 210 (heading - corporate name) was repeated: the kanji form "**国立情報学研究所**" was included in $a without $7; the katakana-yomi form "**コクリツ ジョウホウガク ケンキュウジョ**" was included $a in the field which has "dc" (Japanese - Kana) in $7; and the romaji-yomi form "Kokuritu zyouhougaku kenkyuuzyo" was included in $a in the field where $7 is "ba" (Latin character). Each $6 represented that it had a relation of different character types with the same content ("a") and was a single set (with the same linking number "01").

Heading - Corporate name	210	02	$6	a01
			$a	**国立情報学研究所#**
	210	02	$6	a01
			$7	dc
			$a	**コクリツ ジョウホウガク ケンキュウジョ#**
	210	02	$6	a01
			$7	ba
			$a	Kokuritu zyouhougaku kenkyuuzyo#

Figure 2: A partial example of JAPAN/MARC (A)

4.3 JAPAN/MARC UNIMARC format

Since the second edition of the UNIMARC Manual (1994), it has been possible to record a set of kanji, katakana-yomi and romaji-yomi forms using $6 and $7 in the bibliographic format as well as in UNIMARC/Authorities. However, the NDL did not make any major changes to the JAPAN/MARC bibliographic format for domestic distribution of bibliographic data. On the other hand, the JAPAN/MARC UNIMARC version was developed for international distribution and started to be provided in January 2003.

In the JAPAN/MARC UNIMARC format, the fields specified in the UNIMARC format were used in principle. For example, title and responsibility were recorded in field 200 instead of 251. As in JAPAN/MARC (A), the character type codes were recorded in $7 for katakana-yomi and romaji-yomi forms, and they were recorded as a single set with kanji forms without $7.

Figure 3 shows that field 225 (series title) was repeated, and each form was recorded in $a: kanji form "**物理学**" without $7, katakana-yomi form "**ブツリガク**" with "dc" (Japanese - Kana) in $7, and romaji-yomi form "Buturigaku" with "ba" (Latin character) in $7. Each $6 represents that it has a relation of different character types with the same content ("a" in $6) and one set (with the same linking number "01" in $6). Figure 3 also shows that "**一般相対性理論**", "**イッパン ソウタイセイ リロン**", and "Ippan soutaisei riron" are one set (with the same linking number "02" in $6).

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

Series title	225	1#	$6	a01
			$a	**物理学**
	225	1#	$6	a01
			$7	dc
			$a	**ブツリガク**
	225	1#	$6	a01
			$7	ba
			$a	Buturigaku
	225	1#	$6	a02
			$a	**一般相対性理論**
	225	1#	$6	a02
			$7	dc
			$a	**イッパン　ソウタイセイ　リロン**
	225	1#	$6	a02
			$7	ba
			$a	Ippan soutaisei riron

Figure 3: A partial example of JAPAN/MARC UNIMARC format

The NDL provided bibliographic data to OCLC using this JAPAN/MARC UNIMARC format from November 2010 to December 2011.

4.4 JAPAN/MARC after 2012: JAPAN/MARC MARC 21 format

The NDL implemented the MARC 21 format for both bibliographic and authority data in January 2012. That led to a major change in the JAPAN/MARC format, including the fields for recording readings. The JAPAN/MARC MARC 21 format uses a set of field 880 and other fields (i.e. regular fields) to record kanji form, katakana-yomi form, and romaji-yomi form according to Appendix D "Multiscript Records"[5] of the MARC 21 format. In the JAPAN/MARC MARC21 format, a set of field 880 and regular fields form a set of kanji and its reading forms. In $6, the field number of the set, the occurrence number for matching the field to be linked, and in the case of field 880, the character type code are recorded to represent the set of kanji and reading forms. However, unlike the example given in Appendix D, the kanji form is recorded in the regular field, and the katakana-yomi form and romaji-yomi form are recorded in field 880. In Figure 4, the $6 in field 245 (title statement) is a single set with field 880 that has occurrence number "01", and in field 880, "245-01" in the $6 represents that this field 880 is a single set with the field 245 that has the occurrence number "01", followed by recording the character type codes "$1" (Chinese, Japanese, Korean) and "(B" (Latin). Figure 4 shows the kanji form of 245$a (title), "**理論物理に潜む部分多様体幾何**", the katakana-yomi form of 880$a, "**リロン ブツリ ニ ヒソム ブブン タヨウタイ キカ**", and the romaji-yomi form of 880$a, "Riron butsuri ni hisomu bubun tayotai kika", represent a single set.

[5] https://www.loc.gov/marc/bibliographic/ecbdmulti.html

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

Title Statement	245	00	$6	880-01
			$a	理論物理に潜む部分多様体幾何 =
			$b	Submanifold Geometry Hidden in Theoretical Physics : 一般相対性理論・ゲージ理論との関わり /
			$c	小池直之 著
			$0	001285136
Alternate Graphic Representation	880	00	$6	245-01/$1
			$a	リロン ブツリ ニ ヒソム ブブン タヨウタイ キカ :
			$b	イッパン ソウタイセイ リロン ゲージ リロン トノ カカワリ
Alternate Graphic Representation	880	00	$6	245-01/(B
			$a	Riron butsuri ni hisomu bubun tayotai kika :
			$b	Ippan sotaisei riron geji riron tono kakawari

Figure 4: A partial example of JAPAN/MARC MARC 21 format

Since January 2012, the NDL has been providing bibliographic and authority data to OCLC in the JAPAN/MARC MARC 21 format, instead of the JAPAN/MARC UNIMARC format.

5. Representation of readings other than the MARC format

In addition to the MARC format, the NDL provides bibliographic and authority data in various formats. The NDL Search system provides bibliographic data not only in the MARC format and MARC tag format, but also DC-NDL (RDF) format and JSON format.
The DC-NDL format is a set of metadata description elements and rules based on the Dublin Core (DC), with some additional NDL-specific vocabulary. “dcndl:transcription” is the NDL’s own vocabulary for recording readings. For example, a set of kanji and katakana-yomi of a title can be expressed by recording the kanji form in <rdf:value> and the katakana-yomi form in <dcndl:transcription> in <dc:title>. Neither DC-NDL (RDF) format nor JSON format provides romaji-yomi form.

```
<dcterms:title>理論物理に潜む部分多様体幾何 = Submanifold Geometry Hidden in Theoretical Physics : 一般相対性理論・ゲージ理論との関わり</dcterms:title>
  <dc:title>
    <rdf:Description>
      <rdf:value>理論物理に潜む部分多様体幾何 = Submanifold Geometry Hidden in Theoretical Physics : 一般相対性理論・ゲージ理論との関わり</rdf:value>
      <dcndl:transcription>リロン ブツリ ニ ヒソム ブブン タヨウタイ キカ : イッパン ソウタイセイ リロン ゲージ リロン トノ カカワリ</dcndl:transcription>
    </rdf:Description>
  </dc:title>
```

Figure 5: A partial example of DC-NDL (RDF) format

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450


```
title:
    0:
        value:          "理論物理に潜む部分多様体幾何 = Submanifold Geometry Hidden in Theoretical Physics : 一般相対性理論・ゲージ理論との関わり"
        transcription:  "リロン ブツリ ニ ヒソム ブブン タヨウタイ キカ : イッパン ソウタイセイリロン ゲージ リロン トノ カカワリ"
```

Figure 6: A partial example of JSON format

The Web NDL Authorities service provides authority data in RDF/XML format, RDF/Turtle format, and JSON-LD format. In the Web NDL Authorities, <dcndl:transcription> is used in RDF/XML format as well as in the NDL Search, however, it is used with language attributes so that romaji-yomi form can be expressed in addition to katakana-yomi forms. For example, in <skosxl:prefLabel>, the kanji form is recorded in <xl:literalForm>, the katakana-yomi form is recorded in <dcndl:transcription> with xml:lang="ja-Kana", and the romaji-yomi form with xml:lang="ja-Latn". Turtle format and JSON-LD format also use language attributes to express kanji form and katakana-yomi form as a single set.

```
<xl:prefLabel>
        <rdf:Description>
            <xl:literalForm>小池, 直之</xl:literalForm>
            <ndl:transcription xml:lang="ja-Kana">コイケ, ナオユキ</ndl:transcription>
            <ndl:transcription xml:lang="ja-Latn">Koike, Naoyuki</ndl:transcription>
        </rdf:Description>
    </xl:prefLabel>
```

Figure 7: A partial example of RDF/XML format in Web NDL Authorities

6. Japanese "yomi" in OCLC's WorldCat

As mentioned above, the NDL has been providing JAPAN/MARC to OCLC since November 2010. The NDL has also been providing JAPAN/MARC (A) to VIAF, Virtual International Authority File, since October 2012.
In OCLC's WorldCat, the kanji and romaji-yomi forms are displayed together for the title, author, and publisher, as shown in Figure 8.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

Figure 8: An example of WorldCat's display. ©2022 OCLC, Inc. This Screenshot is used with OCLC's permission. WorldCat is a registered trademark/service mark of OCLC.

Since WorldCat receives data from various organizations around the world, there can be many variations in the romanization of Japanese. The variations in long vowel expressions as described in Section 3 can be found in WorldCat. For example, the romanization of the title "**重要文化財慶應義塾図書館保存修理工事報告書**" in Figure 8 is indicated as "Jūyō bunkazai Keiō gijuku toshokan hozon shūri kōji hōkokusho" with macrons (¯). On the other hand, the bibliographic data in the JAPAN/MARC MARC 21 format represents the romaji-yomi form without any long vowel symbols, "Juyo bunkazai keio gijuku toshokan hozon shuri koji hokokusho" (Figure 9).

880	00	$6	245-01/$1
		$a	ジュウヨウ ブンカザイ ケイオウ ギジュク トショカン ホゾン シュウリ コウジ ホウコクショ.
		$n	ホンペン
880	00	$6	245-01/(B
		$a	Juyo bunkazai keio gijuku toshokan hozon shuri koji hokokusho.
		$n	Honpen

Figure 9: A partial example of JAPAN/MARC MARC21 format

Also, searching the word given as an example in Section 3, "**お母さん**," "okaasan," and "okasan", the search results are different for each. As of January 17, 2022, a keyword search on WorldCat for "**お母さん**" yields about 4,671 results, "okaasan" yields about 2,418 results, and "okasan" yields about 4,171 results.

To help deal with the differences in romaji-yomi expression, it would be desirable to be able to handle kanji and its readings as sets in overseas bibliographic data as well.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

7. A study on the representation of Japanese "yomi" in BIBFRAME

The "BIBFRAME 2.0 to MARC 21 Conversion Specifications"[6] published by the Library of Congress in 2020 uses the MARC 21's Appendix D, Model B without transliterations for multilingual notation. In the "MARC 21 to BIBFRAME 2.0 Conversion Specifications"[7] published in 2021, only for title statement (field 245), paired non-Latin characters and transliterations are converted using xml language codes to maintain correspondence[8]. The title "**重要文化財慶應義塾図書館保存修理工事報告書**" will be represented as shown in Figure 10.

```
<bf:title>
  <bf:Title>
    <bf:mainTitle xml:lang="ja-Latn">Juyo bunkazai keio gijuku toshokan hozon shuri koji hokokusho. Honpen</bf:mainTitle>
    <bf:mainTitle xml:lang="ja-Latn">ジュウヨウ ブンカザイ ケイオウ ギジュク トショカン ホゾン シュウリ コウジ ホウコクショ. ホンペン</bf:mainTitle>
    <bf:mainTitle xml:lang="ja-Jpan">重要文化財慶應義塾図書館保存修理工事報告書. 本編</bf:mainTitle>
    <rdfs:label>重要文化財慶應義塾図書館保存修理工事報告書. 本編</rdfs:label>
  </bf:Title>
</bf:title>
```

Figure 10: A predicted partial example of BIBFRAME in RDF/XML format

8. Next steps

As mentioned above, both the kanji form and its reading form are necessary for identification of Japanese entities. The NDL has been providing sets of the kanji, katakana-yomi and romaji-yomi forms in the bibliographic data and authority data of Japanese materials.

The "Plan for Creating and Providing the National Diet Library's Bibliographic Data 2021–2025," a five-year plan for bibliographic data formulated by the NDL in March 2021, has supporting the new bibliographic framework as one of its items. The NDL plans to study the possibility of providing bibliographic and authority data based on BIBFRAME, which is becoming popular overseas as a new bibliographic framework that can realize the functions of linked data.

The conversion proposed by the Library of Congress only converts non-Latin characters, except for some fields such as the main title. Since the NDL needs to provide both the kanji form and its reading form not only for the title proper but also for parallel titles, other title information, etc., it is necessary to examine the specifications needed to represent the kanji form, katakana-yomi forms, and romaji-yomi forms as a single set, including the use of language codes or script codes, or using original vocabulary such as <dcndl:transcription>.

[6] https://www.loc.gov/bibframe/bftm/

[7] https://www.loc.gov/bibframe/mtbf/

[8] "MARC 21 to BIBFRAME 2.0 Conversion Specifications Ver,1.7" (2021-10-21) is available in GitHub. It allowed converting fields 250, 26X, and 490 as well as 245, keeping the correspondence between non-Latin characters and transliterated forms. https://github.com/lcnetdev/marc2bibframe2/tree/master/spec

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

It is necessary to provide bibliographic and authority data to many more users by supporting the new bibliographic framework, while continuing to support existing JAPAN/MARC users. No matter what bibliographic framework is used, it is necessary to be able to record a set of kanji, katakana-yomi, and romaji-yomi forms for Japanese bibliographic and authority data. The NDL will continue to study the possibility of providing bibliographic data based on BIBFRAME.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

References

Agenbroad, James E. BA, and MLS. 2006. "Romanization Is Not Enough." *Cataloging & Classification Quarterly* 42:2, 21-34. Accessed March 9, 2022. doi: 10.1300/J104v42n02_0.

Aihara, Nobuya, and Naito, Eisuke. 1997. "CJK(中国語/日本語/韓国語)環境下での国際標準件名標目システムの適用 = Adaptation of Universal Standardized Subject Heading System in the CJK (Chinese / Japanese / Korean) Environment." 学術情報センター紀要 = *Research bulletin of the National Center for Science Information Systems* 9:73-81. Accessed March 9, 2022. https://www.nii.ac.jp/journal/bulletins/pdf/bulletin_9.pdf.

Barešová, Ivona. 2016. *Japanese given names : a window into contemporary Japanese society First edition.* Olomouc : Palacký University Olomouc (Monographs).

El-Sherbini, Magda, & Chen, Sherab. 2011. "An Assessment of the Need to Provide Non-Roman Subject Access to the Library Online Catalog." *Cataloging & Classification Quarterly* 49:6, 457-483. Accessed March 9, 2022. doi: 10.1080/01639374.2011.603108.

Hadamitzky, Wolfgang, and Spahn, Mark. [2011]. *Japanese Kanji and Kana : a guide to the Japanese writing system.* Tokyo : Tuttle Publishing.

Hasegawa, Yoko. 2015. *Japanese : a linguistic introduction.* Cambridge : Cambridge University Press.

Edited by Hopkinson, Alan. 2008. *UNIMARC Manual Bibliographic Format. 3rd edition.* München: Saur (IFLA Series in Bibliographic Control, 36).

Iwasaki, Shoichi. [2013]. *Japanese. Revised edition.* Amsterdam : John Benjamins Publishing Company (London Oriental and African language library ; volume 17).

KAMEI, Haruko. 1997. "欧米のCJKシステム : OCLCを中心に = China and Chinese character (Kanji) : to know about Asia 2. CJK systems in the United States: OCLC CJK." 情報の科学と技術 = *The journal of Information Science and Technology Association* 47:7, 334-339. Accessed March 9, 2022. https://doi.org/10.18919/jkg.47.7_334.

Kimura, Maiko. 2015. "A Comparison of Recorded Authority Data Elements and the RDA Framework in Chinese Character Cultures." *Cataloging & Classification Quarterly* 53:7, 753-784. Accessed March 9, 2022. doi: 10.1080/01639374.2014.977984.

Kudo, Yoko. 2011. "Modified Hepburn Romanization System in Japanese Language Cataloging: Where to Look, What to Follow." *Cataloging & Classification Quarterly* 49:2, 97-120. Accessed March 9, 2022. doi: 10.1080/01639374.2011.536751.

LI Chenying ; XIE Qinfang ; YU Shuangshuang ; HAN Mingjie. 2006. "CALIS日本語書誌作成と典拠コントロールの全面的実現ーCALIS共同目録のオンライン典拠コントロールシステムについてー = Implementation of Japanese Literature Cataloging and Authority Control : The Online Authority Control System for the CALIS Union Catalog." 大学図書館研究 = *Journal of college and university libraries* 77, 51-59. Accessed March 9, 2022. https://doi.org/10.20722/jcul.1218.

JLIS.it vol. 13, no. 2 (May 2022)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-450

MATSUI Junko. 2018. "書誌コントロールと図書館目録 = Bibliographic control and OPAC: online public access catalog." 図書館界 = *The library world* 70:1, 287-304. Accessed March 9, 2022. https://doi.org/10.20628/toshokankai.70.1_287.

Naito, Eisuke. 2004. "Names of the Far East: Japanese, Chinese, and Korean Authority Control." *Cataloging & Classification Quarterly* 38:3-4, 251-268. Accessed March 9, 2022. doi: 10.1300/J104v38n03_19.

NAKAI, Machiko. 2007. "日本の全国書誌サービス—その歩みと展望 = National bibliographic service in Japan—History and prospects." 情報管理 = *Journal of information processing and management* 50:4, 193-200. Accessed March 9, 2022. https://doi.org/10.1241/johokanri.50.193.

National Diet Library. 2003. *JAPAN/MARC Manual. Authorities. First edition.* Tokyo: National Diet Library.

National Diet Library. 2002. *JAPAN/MARC Manual. Monographs & Serials. First edition.* Tokyo: National Diet Library.

Edited by Plassard, Marie-France. 2001. *UNIMARC Manual : authorities format. 2nd revised and enlarged edition.* München: Saur (UBCIM Publications ; New Series, Vol. 22).

Takawashi, Tadayoshi BE. 2002. "Cataloging in Japan: Relationship Between Japanese and Western Cataloging Rules." *Cataloging & Classification Quarterly* 35:1-2, 209-225. Accessed March 9, 2022. doi: 10.1300/J104v35n01_13.

TATUOKA, Hirosi, and YAMADA, Hisao. 1996. "On the Theory and Practice of the Romanized Writing System for Japanese = 日本語のローマ字書き体系の理論と実状について." 学術情報センター紀要 = *Research bulletin of the National Center for Science Information Systems* 8:27-74. Accessed March 9, 2022. https://www.nii.ac.jp/journal/bulletins/pdf/bulletin_8.pdf.

Taylor, Insup, Taylor, M. Martin. [2014]. *Writing and literacy in Chinese, Korean and Japanese. Revised edition.* Amsterdam : John Benjamins Publishing Company (Studies in written language and literacy ; volume 14).

Tokita, Takuya, Kimura, Maiko, Miyata, Yosuke, Yokoyama, Yukio, Taniguchi, Shoichi, & Ueda, Shuichi. 2012. "Identifying Works of Japanese Classics for Construction of FRBRized OPACs." *Cataloging & Classification Quarterly* 50:5-7, 670-687. Accessed March 9, 2022. doi:10.1080/01639374.2012.681605.

Yanaike, Makoto. 2009. "現代日本語の音・訓読み分けの機構を論じ、「漢語・和語形態素の相補的分布」に及ぶ = The mechanism of the discrimination between On-yomi and Kun-yomi of Kanji in present-day Japanese." 藝文研究 = The geibun-kenkyu : journal of arts and letters 96:75(202)-95(182). Accessed March 9, 2022. https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00072643-00960001-0095.