

Title	JAPAN/MARC書誌レコードからNCR2018エレメントベースのメタデータへの変換
Sub Title	Conversion of JAPAN/MARC bibliographic records to NCR2018 element-based metadata
Author	谷口, 祥一(Taniguchi, Shoichi)
Publisher	三田図書館・情報学会
Publication year	2022
Jtitle	Library and information science No.88 (2022. ) ,p.49- 71
JaLC DOI	
Abstract	<p>【目的】国立国会図書館により日本目録規則2018年版（NCR2018）を適用し作成されたJAPAN/MARC（JPMARC）書誌レコードに対して，NCR2018の構成に沿ったメタデータすなわち実体とそのエレメントからなるメタデータに再構成すること，そのために必要なスキーマレベルのマッピングの策定とそれを活用したツールを開発し実行することを目的とする。これによって，広義の学習用コンテンツとして利用可能なNCR2018のデータ事例を大幅に増やすことができる。</p> <p>【方法】先行事例を参照しつつも，独自にマッピングを策定した。マッピングはJPMARCデータを変換し再構成することに目的を限定し，かつJPMARCフォーマットで表現されている要素は基本的にすべてマッピング対象とした。こうした方針の下，マッピングにおける個々の課題に対して適切な対応策を導入しそれらの解決を図った。さらに，策定したマッピングに基づいてデータ変換を実行するツールを実装し，変換を試行した。</p> <p>【結果】策定したマッピングおよび変換用ツールは，1）マッピングとその後の調整処理を切り分ける，2）マッピングにおいてNCRエレメントに適宜，限定子を導入し，その後の調整処理において活用する，3）JPMARCデータ要素単位は「フィールド識別子，インディケータ，サブフィールド識別子」の組み合わせとし，場合によってはさらに「資料区分」を組み合わせ，より精度の高いマッピングとする，4）優先度を設けて2レベルに分けたマッピングとする，5）個々の値の由来情報を併せて記録することなどによって，課題群に対処した。併せて，策定したマッピングについて，カーディナリティのパターンに分け，詳細にその該当例の検討や残された課題の整理などを行い，適切なマッピングが困難な部分を同定した。また，実装したツールによってデータ変換を実行し，87,000件のレコードを変換し，NCR2018の構成に沿ったメタデータを作成することができた。</p> <p>Purpose: For reconstructing the JAPAN/MARC (JPMARC) bibliographic records created by the National Diet Library in accordance with the Nippon Cataloging Rules, 2018 ed. (NCR2018) into metadata that is composed of NCR2018 entities and their elements, the purpose of this study is to develop the necessary schema-level mapping and tools that make use of it, and then to perform data conversion. This will contribute to a considerable increase of the NCR2018 data examples available as learning content.</p> <p>Methods: We developed our own mapping to convert and reconstruct the JPMARC data while basically considering all the data elements in the JPMARC format for mapping. Based on this policy, appropriate methods were introduced to solve the individual mapping issues. In addition, we implemented a tool to perform data conversion based on the mapping.</p> <p>Results: The characteristics of the mapping and conversion tool are as follows: 1) separate the mapping from the subsequent adjustment process, 2) introduce qualifiers to NCR elements in the mapping and utilize them in the subsequent adjustment process, 3) adopt the JPMARC data element unit as a combination of</p>

	“field tag, indicator, and subfield code” and in some cases “material designation” is also combined for more accurate mapping, 4) mapping is divided into two levels with priority, and 5) the provenance information of each value is recorded. Furthermore, the mapping was divided into patterns of cardinality, and a detailed examination of the remaining issues was conducted. The implemented tool performed data conversion of 87,000 records, thereby resulting in a vast amount of metadata in line with the NCR2018 structure.
Notes	原著論文
Genre	Journal Article
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000088-0049">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000088-0049</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

## 原著論文

# JAPAN/MARC 書誌レコードから NCR2018 エlementベースの メタデータへの変換

## Conversion of JAPAN/MARC Bibliographic Records to NCR2018 Element-based Metadata

谷口 祥一

*Shoichi TANIGUCHI*

### *Résumé*

**Purpose:** For reconstructing the JAPAN/MARC (JPMARC) bibliographic records created by the National Diet Library in accordance with the Nippon Cataloging Rules, 2018 ed. (NCR2018) into metadata that is composed of NCR2018 entities and their elements, the purpose of this study is to develop the necessary schema-level mapping and tools that make use of it, and then to perform data conversion. This will contribute to a considerable increase of the NCR2018 data examples available as learning content.

**Methods:** We developed our own mapping to convert and reconstruct the JPMARC data while basically considering all the data elements in the JPMARC format for mapping. Based on this policy, appropriate methods were introduced to solve the individual mapping issues. In addition, we implemented a tool to perform data conversion based on the mapping.

**Results:** The characteristics of the mapping and conversion tool are as follows: 1) separate the mapping from the subsequent adjustment process, 2) introduce qualifiers to NCR elements in the mapping and utilize them in the subsequent adjustment process, 3) adopt the JPMARC data element unit as a combination of “field tag, indicator, and subfield code” and in some cases “material designation” is also combined for more accurate mapping, 4) mapping is divided into two levels with priority, and 5) the provenance information of each value is recorded. Furthermore, the mapping was divided into patterns of cardinality, and a detailed examination of the remaining issues was conducted. The implemented tool performed data conversion of 87,000 records, thereby resulting in a vast amount of metadata in line with the NCR2018 structure.

---

谷口祥一：慶應義塾大学文学部，〒108-8345 東京都港区三田 2-15-45

Shoichi TANIGUCHI: School of Library and Information Science, Keio University, 2-15-45 Mita, Minatoku, Tokyo 108-8345

e-mail: taniguchi@z2.keio.jp

受付日：2022年8月4日 改訂稿受付日：2022年10月14日 受理日：2022年10月25日

発行日：2022年12月30日

DOI: 10.46895/lis.88.49

- I. はじめに
    - A. 研究背景と研究目的
    - B. 先行研究と事例
  - II. マッピングの検討と策定
    - A. メタデータマッピングの問題点
    - B. 本研究におけるマッピング方針
    - C. マッピング元とマッピング先の単位設定
    - D. 策定したマッピングテーブル
  - III. 策定したマッピングの概要と課題整理
    - A. 由来情報の記載, 「読み」と「情報源」のマッピング
    - B. 「1対1」のマッピング
    - C. 「1対多」のマッピング
    - D. 「多対1」のマッピング
    - E. 「多対多」のマッピング
    - F. マッピング困難なデータ要素
  - IV. 策定したマッピングに基づく JAPAN/MARC レコードの変換処理
  - V. おわりに
- 注・引用文献

## I. はじめに

### A. 研究背景と研究目的

日本目録規則 2018 年版（以下、NCR2018 または NCR）<sup>1)</sup> は、従来の 1987 年版（NCR1987）と比べ内容や構成が大きく異なる。それは、書誌レコードの機能要件（FRBR）等の明確な概念モデルへの依拠、資料の内容的側面と物理的側面の整理、意味的側面と構文的側面の分離、2018 年当時の Resource Description and Access（RDA）との相互運用性の担保などに起因する変更である。国立国会図書館（NDL）による国内刊行資料に対する NCR2018 の適用（2021 年 1 月から）、図書館流通センター（TRC）による適用（2022 年 1 月から）が既に開始されており、今後はより多くの機関による採用と適用済みデータの公開や提供などによって、標準的に適用される記述規則となることが確実視される。

しかしながら、旧来の記述規則と大きく考え方や構成が変更されたものでありながら、現時点ではその理解に資するデータ事例、特に広義の学習・教育・訓練用のデータ例は極めて限定され

る。NCR2018 の構成に沿った現時点のデータ事例は、1) 日本図書館協会目録委員会（JLA 目録委員会）がウェブページから公開している事例群<sup>2)</sup>、2) 木村麻衣子による解説書<sup>3)</sup>、3) 鳥海恵司による主に音楽作品とその資料群に対するデータ実例集<sup>4,5)</sup>、4) 司書課程用の一部の教科書の記載例<sup>6)</sup> などにとどまる。鳥海によるデータ実例集を除いては、いずれも事例数は限られている。

他方、前述した NDL や TRC によって NCR2018 を適用し作成された書誌・典拠レコードは、既に相当数が存在する。これらも学習用の事例データとして参照し活用できる有効なデータと考えられるが、その記録フォーマット（スキーマ）はいずれも旧来からの MARC フォーマットを微調整したものであり、NCR2018 によって作成されるメタデータをそのまま素直に表現したものではない。NDL は MARC21 を採用し、TRC は UNIMARC ベースの以前の JAPAN/MARC フォーマットに準拠したものを採用しており、それらのデータをそのまま見ただけでは NCR2018（そしてその基盤にある FRBR）によって、どのように対象とする情報資源やその形成する事象が捉えられ表現さ

れたのかは極めて分かりにくい。

以前のNCR1987や英米目録規則第2版とMARCフォーマットとは相互運用性が相当程度に担保されていた。すなわち、NCR1987に依拠して作成されるデータを、構成上の大きな変換を伴わずに、そのままMARCフォーマットに収容することができた。両者が基づく概念モデルが曖昧なものであっても共通していたということである。それに対して、NCR2018やRDAとMARCフォーマットの間には、大きな構造的相違が存在する。前者はFRBRやIFLA図書館参照モデル(IFLA LRM)における実体とその属性や関連からなる構造を基本とするのに対して、後者のMARCフォーマットは旧来からの構成を、微調整を加えつつ引き続き保持している。それゆえ、MARCフォーマットに記録されたデータを見ても、NCR2018やRDAによる捉え方や構成がそのままでは極めて分かりにくい。

本研究の目的は、NDLによりNCR2018を適用し作成されたMARCデータに対して、NCR2018の構成に沿ったメタデータ、すなわち実体とそのエレメントからなるメタデータに再構成すること、そのために必要なスキーマレベルのマッピングの策定とそれを活用したツールを開発し実行することである。これによって、メタデータ作成者(カタログガー)がNCR2018に従ってどのように対象とする事象を捉え、それをメタデータとして表現しているのかを、理解が容易な形式で表示する。これは広義の学習用コンテンツとなることを意図しており、現時点でのそれらの不足を補うものと位置づける。

ただし、あくまでもNDLによるNCR2018適用細則<sup>7)</sup>に基づくデータ事例であり、多数の別法や任意規定を備えたNCR2018における一つの適用例に限定される。それゆえ、NDLが現時点で採用しているNCR実体とエレメントの範囲の記述に限定される。NCR2018が規定する最大限のエレメントを用いて対象事象を表現したデータではないという制約がある。また、今回はMARC書誌レコードのみ対象とする。

上記目的には、2つの下位目的が包含されて

いる。1つは、JAPAN/MARC MARC21フォーマット(以下、JPMARCまたはMARC)<sup>8)</sup>とNCR2018の実体およびエレメントとのスキーマレベルのマッピングを検討し策定することである。こうしたマッピングにどのような問題が潜んでいるのか、具体的に検証し、可能な対処策を検討する。併せて、NCR2018に適合しないデータ要素の同定を図る。2つめは、策定したマッピングおよび対処策に基づき、実際にJPMARCデータを変換するツールを開発し実行することである。こうしたツールの開発と実行によって、NCR2018に従った形式に再構成した、かつ相当数の学習用データを提供することができる。他方、変換処理結果の学習用コンテンツとしての有効性や学習者による評価など、その評価に関わる部分は本研究では扱わない。変換処理の対象とするデータ項目の範囲および値の変換処理内容も具体的な利用目的・利用方法に依存してその妥当性が決まる部分があるが、本研究ではその点を限定せず、広く学習用に向けてJPMARCに記録されているすべてのデータ項目およびその値を最大限活用する方針とした。こうした点も含めて、有効性の評価については別途の検討が必要になる。

## B. 先行研究と事例

本研究の直接的な先行事例には、RDAの維持管理を担うRDA Steering Committee (RSC)がRDAに関連するツールとして公開しているRIMMF (RDA in Many Metadata Formats)<sup>9)</sup>がある。現在、バージョン3 (RIMMF3)と4があり、前者のバージョン3は3Rプロジェクト(2020年末完了)以前のRDA、すなわちNCR2018策定時に参照していたRDAに概ね対応し、現在開発途中のバージョン4は3Rプロジェクト後のRDAに対応する。当該ツールはRDAによるメタデータ作成作業や学習など、幅広く活用できる高機能なツールである<sup>10, 11)</sup>。RDAの実体とエレメントに沿ってメタデータを作成できることは勿論、MARC21のレコードを読み込みRDAに従った構成に変換する機能を備えており、そうしたデータに手修正を加えることも可能である。

当該ツールを使用すれば、MARCXML 形式に変換後の JPMARC レコードを読み込み、それを RDA の実体とそのエレメントによる表示に変換することができる。さらに、設定ファイルを修正することにより、RDA エレメントのラベルを NCR2018 によるものに置換することができる。しかしながら、それを超えての機能追加や変更を加えることは困難である。NCR2018 および JPMARC に特有の事項には対処できないという限界がある。

なお、同ツールにおいて採用されている MARC21 データ要素と RDA エレメントとのマッピングが、インストール後のファイルにおいて参照可能である<sup>12)</sup>。このマッピングは、本研究におけるマッピング策定時にも参考とした（詳細は後述する）。

RIMMF において採用されている上記のマッピングとは別に、3R プロジェクト以前の MARC21 と RDA との異なるマッピングが RSC 自身によって公開されている<sup>13)</sup>。しかし、RIMMF のそれとは方針が異なり、意味的に対応する部分を有すると判断された最大限の MARC データ要素と RDA エレメントを対応づけている。その結果、メタデータを変換するという目的から見た実用性に乏しく、本研究では参考にしていない。同様に、RDA Registry において、最新の RDA から MARC21 へのマッピングとアラインメントが公開されているが<sup>14, 15)</sup>、その設定方針は上記と同じであり、メタデータ変換の目的に適うものではない。

MARC21 のデータ要素すべてに対するマッピングではないが、NCR2018 や RDA が基盤とする概念モデルである FRBR 等の実体のうち、特に著作（場合によっては、それに表現形が加わる）の抽出と照合などが、OPAC の「FRBR 化」(FRBRization) という名の下で研究および試行されていたことがある<sup>16-18)</sup>。その場合、主対象は NCR2018 や RDA 適用以前の旧来の書誌レコード群である。こうした FRBR 化は著作（および表現形）を示すデータ要素を指定し、そこから値を抽出し、著作等のデータを生成することに目的が限定されており、それ以外のデータ要素に

関するマッピングなどは含まれない。これらゆえ、本研究の直接的な先行研究にはなりえない。

ところで、NCR2018 と同程度に MARC21 とは構造的に差異がある RDA (3R プロジェクト後はさらに差異が拡大) に関しても、その構造に沿ったデータ作成事例の公開は限られている。RIMMF によるデータ例があるほか、RSC が公開している少数例の事例があるのみで<sup>19)</sup>、RDA の解説書等においても MARC21 フォーマットによるデータ例のみ記載されている<sup>20)</sup>。欧米ではあくまでも MARC21 による記録が前提とされ、元の RDA の構成に沿った表示などが必要とされないということなのであろうか、この点は疑問である。

MARC フォーマットから他のスキーマへのマッピング、そしてそれに基づく変換例に目を転ずれば、その事例は複数存在する。NDL に限っても JPMARC 書誌レコードから DC-NDL への変換、同典拠レコードの Web NDL Authorities RDF データへの変換などが行われている。これらは元データの主要部分のみの変換を意図しており<sup>21)</sup>、基本的にすべてのデータ要素を変換対象にするといった本研究が目指す包括的な変換を志向したものではない。あるいは、米国議会図書館によって MARC21 と MODS, Dublin Core, ONIX 等との間のマッピングが公開されているが<sup>22)</sup>、これらを含めてマッピングに伴う問題点とその検討などは報告されていない。

## II. マッピングの検討と策定

### A. メタデータマッピングの問題点

メタデータマッピング (metadata mapping) とは、それぞれ独立したメタデータスキーマ間で、通常、等価の（もしくはそれに近い）データ要素間に対応づけを行うことであり、マッピングをテーブル等で表現したものをクロスウォーク (crosswalk) とも呼ぶ<sup>23)</sup>。これらはメタデータの統合的検索や変換などに用いることを目的としており、その利用目的に依存してマッピングの具体的な方針や結果も変わりうる。

メタデータの構成に即していえば、そのデー

タ要素のマッピングと、値の記録に用いる用語のマッピングに分けることができる。RDF (Resource Description Framework) の区分に従えば、前者が RDF プロパティ (とその主語・目的語リソースとなる RDF クラス)、後者が RDF クラスにおけるマッピングとなる。加えて、マッピングはメタデータの構造的・構文的な側面に関わる問題と、意味の側面の問題に分けて捉えることができる。マッピングという問題の全体枠の整理とレビューを行ったものに、Haslhofer によるものがある<sup>24, 25)</sup>。

特に意味の問題が絡むため、具体的なマッピング策定の場面では通常、多様な問題に遭遇する<sup>26)</sup>。マッピングにおける意味的な等価性の問題について、以前から統制語彙に関わる問題として例えば Aitchison らは、1) 厳密な等価 (exact equivalence), 2) 準等価 (inexact equivalence or near-equivalence), 3) 部分等価 (partial equivalence), 4) 一対多の等価 (single to multiple equivalence), 5) 非等価 (non-equivalence) に分けている<sup>27)</sup>。1) はマッピング元とマッピング先との要素 (概念) が等しいこと、2) はマッピング元と先とが部分的に重なりを有するが、相互にはみ出す部分を有すること、3) はマッピング元がマッピング先の要素を包含し、かつ他の要素は包含しないこと、4) はマッピング元が複数のマッピング先を包含していること、5) はマッピング元と先の要素が概念的に重なる部分をもたず相互排他であることを指している。これをマッピングのカーディナリティ (基数) の観点でまとめれば、1) と 3) は「1対1」、4) が「1対多」となる。現実的には、これに「多対1」となるマッピングが加わる。ISO 国際規格 25964-2:2013 も同様な区分を示しており<sup>23)</sup>、さらには同様にマッピングにおける等価性の問題を論じたものは他にも複数ある<sup>28, 29)</sup>。

こうした問題への対処の1つとして、例えば RDF クラス skos:Concept 間に限定されるが、等価の程度を反映したプロパティ skos:exactMatch, skos:closeMatch, skos:broadMatch, skos:narrowMatch, skos:relatedMatch が分けて定義されている<sup>30)</sup>。

このように等価の程度を分けてマッピングを表現するという方策は、本研究においても一部採用する。ただし、こうした方策によって対処できる範囲は限られている。

メタデータマッピングは領域を問わず、これまで多数の実施例があるはずだが、意味の問題への対処を含めた詳細な検討や報告は限られる。

Najjar らは学習用オブジェクトメタデータ (Learning Object Metadata) に関して ARIADNE から IEEE LOM へのマッピングについて比較的詳細に報告している<sup>31)</sup>。大学の機関リポジトリにおけるデジタルオブジェクトのマッピングについて検討した Chau による報告もある<sup>32)</sup>。

## B. 本研究におけるマッピング方針

本研究で策定するマッピングは、JPMARC 書誌レコードフォーマットのデータ要素から NCR2018 の実体およびエレメントへのマッピングである。以下、本研究で採用した方針を記す。

- 1) マッピングの使用目的は、実際の JPMARC データを変換し再構成して表示することである。それゆえ、意味的に最大限の対応関係を示すマッピングとはしない。
- 2) 基本的に実体「表現形」および「個別資料」を主対象とする書誌レコードフォーマットであるが、「著作」および「表現形」に関しても該当するデータ要素を抽出し、その記録を生成する。すなわち、著作の典拠形アクセス・ポイント (AP) を記録する MARC フィールド 730 をもたない場合も、最低限度であれ、著作および表現形を機械的に生成する。また、本研究では典拠データからのマッピングを含めていないため、創作者等となる実体「個人」・「家族」・「団体」へのマッピングについても、該当するデータ要素が出現する範囲内で設定する。
- 3) NDL によって JPMARC フォーマットで表現されている事項は基本的にすべてマッピングの対象とする。ただし、固定長フィールドにおいて、値が一律に固定されているものは除く。こうした意図の下で、NCR エレメントに適切にマッピングできないデータ要素を同定する。

これは NCR2018 に依拠していない箇所の特長でもある。また、現時点の JPMARC において採用されていない、可能なすべての MARC21 データ要素<sup>33)</sup>を対象にして、いわば最大限可能な表現力とその限界などを検討するという立場は取らない。

- 4) マッピング自体とマッピングした後の調整処理を切り分ける。NCR エレメントに依存したマッピング後に必要な調整処理を、マッピングとは別に検討し機能実装を図る。
- 5) 逆方向のマッピングとなる NCR2018 実体およびエレメントに基づくメタデータを、JPMARC を含め MARC21 フォーマットに変換することは意図しない。両方向のマッピングが結果的に同じとなることは多いが、必ずしも同一であるとは限らない。

### C. マッピング元とマッピング先の単位設定

#### 1. 先行事例

JPMARC マニュアルは、フィールド 015 以降のフィールドごとの説明に関わり、サブフィールド識別子ごとに対応する「NCR 条項」（それはそのまま NCR エレメントに対応づけられる）を示している<sup>8)</sup>。フィールド 245（タイトル、責任表示）のサブフィールド a（本タイトル）は NCR 条項「#2.1.1」（本タイトル）に、サブフィールド b（並列タイトル、タイトル関連情報等）は NCR 条項「#2.1.2」（並列タイトル）、「#2.1.3」（タイトル関連情報）、および「#2.1.4」（並列タイトル関連情報）の 3 つにそれぞれ対応づけている。

これらの記載をマッピングとして見た場合、フィールドとサブフィールドを組み合わせたとの NCR エレメントとの双方向のマッピングと受け止められる。ただし、NCR 条項を記載する基準など、そこには一切の説明がないため記載内容から推測する限り、メタデータ変換を意図したものと想定しにくい。意味的に対応するエレメントを広く記載したものと推測される（この点については後述する）。また、対応する NCR 条項の記載がないサブフィールドは、NCR エレメントにマッピングしないという指示と受け止める。

一方、RIMMF3 におけるマッピングは、ツールとして実装し実際にメタデータ変換処理の際に参照するためのものであり、JPMARC マニュアルに比べて限定的なマッピングといえよう。RDA エレメントの単位でそれにマッピングする MARC21 データ要素を示してある。例えば、エレメント title proper（本タイトル）は MARC21 データ要素「245\$.¥anp」からのマッピングと指示されている。同ツールが採用した略記法で記されており、フィールド識別子「245」、第 1 と第 2 インディケータが「..」、すなわちそれぞれを特定せず、任意のインディケータ値を許容することを表している（記号“¥”は区切りを表す）。そしてサブフィールド a, n, p の 3 つが該当するとされている。また、エレメント other title information（タイトル関連情報）は「245\$.¥b¥¥¥」からのマッピングとされており、サブフィールド b の下で区切り記号“.”の箇所が該当するとしている。

これらから、RIMMF3 では可変長フィールドの MARC データ要素については、「フィールド識別子+第 1 インディケータ+第 2 インディケータ+サブフィールド識別子」の組み合わせがマッピングの単位とされ、さらに場合によっては区切り記号を含めた単位を採用していることになる。ちなみに、RSC により別途公開されている MARC データ要素から RDA エレメントへのマッピングテーブルにおいても基本的に上記の単位を採用しているが、そこでは区切り記号による分割は採用していない<sup>13)</sup>。

このように、MARC フォーマットにおけるマッピング単位の設定自体が検討を要する事項である。こうした点の指摘は Thomale によってなされ、そこでは区切り記号による分割の必要性が指摘されている<sup>34)</sup>。また、Coyle は MARC データ要素を管理フィールド (00X)、番号・コード化フィールド (0XX)、記述フィールド (1XX～) に分け、それぞれにおける処理単位を検討している<sup>35)</sup>。その結果、1) 個々のデータ要素が相互に独立して対象リソースを記述する場合、2) あるデータ要素は他のデータ要素値を説明あるいは補足するという連鎖の構成をとる場合、3) 複数の

データ要素が組み合わされて対象リソースを記述する場合という区分を導出している。MARCフォーマットは長期間にわたり使用されてきており、その間に多様な更新が加えられた結果、部分的にせよ、整合的かつ一貫性あるスキーマとは言い難い側面が内包されているといえよう。

## 2. 本研究における採用単位

先行事例等を参照しつつ、本研究におけるマッピング元とマッピング先のデータ単位を決定した。まずNCR2018については、すべてのNCR2018エレメントを対象先の候補とする。エレメントは、多くの場合、エレメント・サブタイプまたはサブエレメントを備えており、マッピングにおいてはできるだけ特定性の高い下位レベルのエレメントに対応づけることが求められる。元の情報からの損失をできるだけ回避するために、適切なレベルにおいて等価となるエレメントに対応づけることを意味する。なお、複数のMARCデータ要素からの多対1となるマッピングにおいて、それぞれのMARCデータ要素の意味範囲を区別することが適切な場合などには、NCRエレメントを擬似的に細分する限定子を導入する（限定子については後述する）。

JPMARCデータ要素については、固定長フィールド（レコードラベル000とフィールド001~008）と可変長フィールド（01X~08X）に分かれ、前者の固定長フィールドについては特定位置のデータ要素値がマッピングの単位となる。それに対して、可変長フィールドは、サブフィールド識別子により区分されるサブフィールドを単位とする。具体的には、「フィールド識別子+第1インディケータ+第2インディケータ+サブフィールド識別子」を組み合わせた単位とする。インディケータの値によってマッピング先を変える場合があるからである。こうした単位設定は、JPMARCマニュアルを除く、MARC21関連のマッピングの先行例と共通する。本研究では区切り記号による分割も、これに加えて部分的に採用した。ただし、十分とはいえない状況にあり、フィールド245など、場合によっては複雑となるパターン

には対応できていない。区切り記号によるデータ要素の識別と分割は安定した方式とはいえず、MARCフォーマットの基本的な問題の1つといえよう。

他方、本研究で独自に追加した方式がある。それはフィールド007の位置00（「資料カテゴリー」。以降は「007/00」という略記法を採用）の値による区分を、場合によっては他フィールドにおける参照単位と組み合わせて使用するという方式である。007/00には10の資料区分がコード化され、文字資料「t」、地図資料「a」、楽譜「q」、録音資料「s」、映像資料「v」、映像資料（静止画）「g」、静止画資料「k」、電子資料「c」、点字資料「f」、マイクロ資料「h」として表現されている<sup>36)</sup>。これを主にフィールド300番台の「キャリアに関する情報」などを適切に仕分けする目的で導入する。

例えば、エレメント「#2.18 大きさ」に対して、下位エレメントである「#2.18.1 地図等の大きさ」と「#2.18.2 静止画の大きさ」がある。前者の#2.18にはJPMARCフィールド300（第1・第2インディケータはいずれも未定義）のサブフィールドc（大きさ）に対応づけることができるが、後者の#2.18.1と#2.18.2を除外するため、上記の資料区分を用いて「c, f, g, h, q, s, t, v」のみ該当することを指定する。それに併せて、#2.18.1と#2.18.2には、それぞれ資料区分「a」と「k」が該当することを指定する。こうした資料区分との組み合わせによって、より精度の高いマッピングとすることができる。フィールド340（媒体の特性等。第1・第2インディケータは未定義）についても、サブフィールドごとに該当する資料区分を指定した上で対応づけるNCRエレメント（「#2.19 基底材」、「#2.21 マウント」など）を決定している。フィールド538のサブフィールドa（システム要件に関する注記）を、資料区分によって対応づけるNCRエレメント（「#2.33 装置・システム要件」、「#2.29.7.1 再生チャンネルの詳細」、「#2.29.8 特定の再生仕様」など）を区別する点も同様である。

D. 策定したマッピングテーブル

本研究で策定したマッピングはテーブルの形式で管理している。第1表にその一部を示した。テーブル全体については、本稿末尾に記載した

URL から参照されたい。

テーブルの左から第1列が「NCR エレメント」、第2列が「エレメント限定子」であり、それにマッピングする JPMARC のデータ要素とし

第1表 策定したマッピングテーブル (部分)

NCR エレメント	エレメント限定子	フィールド 識別子	第1 インディ ケータ	第2 インディ ケータ	サブ フィールド 識別子	資料区分	JPMARC マニ ュ アル	RIMMF3: MARC21 から RDA
#2 体現形								
#2.1 タイトル								
#2.1.1 本タイトル		245	0	0	a		245Ya	245Y.Ya
#2.1.1 本タイトル	巻次等	245	0	0	n		245Yn	245Y.Yn
#2.1.1 本タイトル	部編等のタイトル	245	0	0	p		245Yp	245Y.Yp
#2.1.1 本タイトル	読み	245	0	0	A			
#2.1.1 本タイトル	巻次等読み	245	0	0	N			
#2.1.1 本タイトル	部編等のタイトル 読み	245	0	0	P			
#2.1.1 本タイトル	2番目以降の個別 タイトル等	245	0	0	t		245Yt	
#2.1.1 本タイトル	2番目以降の個別 タイトル等読み	245	0	0	T			
#2.1.2 並列タイトル		245	0	0	b		245Yb	
#2.1.2 並列タイトル		246	0,2,3	1	a		246Ya	290Y.1Ya
#2.1.2 並列タイトル	部編番号	246	0,2,3	1	n			
#2.1.2 並列タイトル	部編名	246	0,2,3	1	p			
#2.1.2 並列タイトル	読み	245	0	0	B			
#2.1.3 タイトル関連情報		245	0	0	b		245Yb	245Y.Yb
#2.1.3 タイトル関連情報	読み	245	0	0	B			
#2.1.4 並列タイトル 関連情報		246	0,2,3	1	b		245Yb, 246Yb	292Y.1Ya
#2.1.4 並列タイトル 関連情報	説明句	246	0,2,3	1	i			
#2.1.4 並列タイトル 関連情報	読み	246	0,2,3	1	A			
#2.13 刊行頻度		008/18						
#2.13 刊行頻度		310	#	#	a		310Ya	310Y.Ya
#2.15 機器種別		337	#	#	a		337Ya	337Y.Ya
#2.15 機器種別	情報源	337	#	#	2			
#2.16 キャリア種別		338	#	#	a		338Ya	338Y.Ya
#2.16 キャリア種別	情報源	338	#	#	2			
#2.17 数量		300	#	#	a		300Yae	300Y.Ya
#2.18 大きさ		007/04				h		
#2.18 大きさ		300	#	#	c	c, f, g, h, k, q, s, t, v	300Yce	300Y.Yc
#2.18 大きさ*		500	#	#	a	c, f, g, h, k, q, s, t, v		
#2.18.1 地図等の大きさ		300	#	#	c	a		300Y.Yc
#2.18.2 静止画の大きさ		300	#	#	c	k		300Y.Yc
#2.19 基底材		340	#	#	a	a, c, f, g, h, k, q, s, t, v	340Ya	340Y.Ya
#2.19 基底材	情報源	340	#	#	2	a, c, f, g, h, k, q, s, t, v		
#2.21 マウント		340	#	#	e	a, c, f, g, h, k, q, s, t, v	340Ye	340Y.Ye
#2.21 マウント	情報源	340	#	#	2	a, c, f, g, h, k, q, s, t, v		
#2.22 制作手段		300	#	#	b	a	300Yb	
#2.22 制作手段		340	#	#	d	c, f, g, h, k, q, s, t, v	340Yd	340Y.Yd
#2.22 制作手段	情報源	340	#	#	2	c, f, g, h, k, q, s, t, v		

\*: マッピングの優先度が低い NCR エレメント

て、フィールド識別子、第1・第2インディケータ、サブフィールド識別子を第3列から第6列に記載している。例えば、「#2.1.1本タイトル」には、「フィールド245、第1・第2インディケータとも0、サブフィールドa」の値をマッピングするという指示である。つまり、第3列から第6列の組み合わせに該当する値を、第1列に示したNCRエレメントにマッピングする。第3列から第6列の組み合わせを、RIMMF3による略記法にならえば上記は245¥00¥aと表記される（以降は字数節約のため、この略記法を採用する）。ここでは資料区分による限定はないため、第7列「資料区分」の欄は空値である。なお、インディケータの欄における“#”は空白を表し、未定義などを指す。マッピングにおいて同じ扱いとする複数のインディケータ値を、テーブル内では並べて記載しているが（例：「0, 2, 3」）、これらは個別のインディケータ値に分解した上で他の要素との組み合わせを生成する。また、サブフィールド内での区切り記号の適用については、マッピングテーブルに含めた記載とはしていない。

策定したマッピングテーブルを集計した結果を第2表に示す。マッピングペア（JPMARCデータ要素とNCRエレメントの組み合わせ）の数は、合計787であった。NCRエレメント異なり数が185、JPMARCデータ要素の異なり数が588となった。この集計では資料区分の組み合わせは捨象している。

次に、資料区分の組み合わせ適用以外に本研究で新たに導入した方策を、以下に記す。

1) エレメント限定子の導入：前述したNCRエレメント「#2.1.1本タイトル」に対しては、先に示した245¥00¥a以外にも、245¥00¥n（巻次または部編番号）や245¥00¥p（部編等のタイトル）など複数の要素がマッピングされ、それぞれを限定子「巻次等」や「部編等のタイトル」をもって区別している（第1表）。巻次等や部編等のタイトルはその意味する範囲において、NCR2018による本タイトルに包含される。このように、NCRエレメントに比してMARCデータ要素の意味範囲が狭く、より限定される場合などに限定子を使用する。あるいは、値の表現法などがコード表現に限定される場合にも、限定子を用いてそれを明示する。

こうした限定子を導入することによって、a) マッピングされた値の意味や位置づけが容易に理解可能となり、またb) 限定子の参照によって、マッピング後の適切な調整処理を実現できる。単一エレメントに複数のJPMARCデータ要素がマッピングされたとき、それらを相互に区別する、あるいはそれらを組み合わせで複合値を形成する、優先すべき値を判別するなどの調整処理において参照することができる。787のマッピングペアのうち、限定子をもたないNCRエレメントは278、限定子をもつもの（エレメントと限定子の組み合わせ）が527となり、数の上では限定子

第2表 マッピング結果の集計

	総数	1対1	1対多	多対1	多対多
マッピングペア数	787	19	298	687	217
NCRエレメント限定子なし	278	18	162	176	86
限定子あり	527	1	136	511	131
NCRエレメント異なり数	185	19	139	85	58
JPMARCデータ要素異なり数	588	19	99	560	90
優先度が低いものを除外					
マッピングペア数	690	26	166	649	151
NCRエレメント限定子なし	179	24	38	143	26
限定子あり	511	2	128	506	125
NCRエレメント異なり数	113	26	51	72	36
JPMARCデータ要素異なり数	588	26	64	558	60

をもつペアが2倍近い結果となった(第2表)。

2) マッピング優先度の適用: JPMARC データ要素の意味において対応づけが可能な複数の NCR エレメントが存在する場合、より等価性が高いエレメントに優先的に対応づける。そして、それ以外のエレメントには記号 “\*” を付して、その優先度を下げることが明示する。例えば、500¥##¥a (一般注記) は幅広い記載内容が想定されるため、NCR エレメント「#2.41 体現形に関する注記」へのマッピングを優先的に設定し、その下位の「#2.41.1 タイトルに関する注記」から「#2.42.3 キャリアに関するその他の情報の変化に関する注記」まで、加えて「#2.18 大きさ」などへは優先度を下げてマッピングする(第1表)。他の事例として、データ要素 245¥00¥c は、その意味に基づき「#2.2.1 本タイトルに関する責任表示」および「#2.2.2 本タイトルに関する並列責任表示」にマッピングできるが、実際に後者に該当する事例は限られているため、その優先度を下げ記号 “\*” を付けてマッピングしている。同じく、MARC データ要素 250¥##¥a は、「#2.3.1 版次」に加えて「#2.3.5 付加的版次」に該当する値を記録している場合もあるが、そうした事例は限られるため、後者はマッピングの優先度を下げる。さらにその可能性の低さから、「#2.3.2 並列版次」と「#2.3.6 並列付加的版次」には、上記データ要素からのマッピングを設定しない。

全体で 787 ペアあったマッピングが、優先度が低いとしたペアを除外したときには、690 ペアに減少する(第2表)。そして、それら減少分の多くが NCR エレメントの限定子がないものであったこと(278 から 179 ペアへの減少)が分かる。

一方、固定長フィールドに関しては、示したテーブルでは簡略にそのマッピングを記している。第1表に示した範囲では、MARC データ要素 008/18 が NCR エレメント「#2.13 刊行頻度」に、007/04 かつ資料区分 h が「#2.18 大きさ」にそれぞれマッピングすることを示している。これらの値はコード化されているため、その意味する内容が分かるよう表示ラベルに置換するテーブルを別途作成した。また、固定長フィールドにおいても、資

料区分に依拠して解釈すべき場合もあり、そのときには該当する資料区分を併せて指定した。

また、策定したマッピングテーブルには、JPMARC マニュアルが示すマッピングおよび RIMMF3 が採用するマッピングを簡略に記し比較可能とした(第1表)。

1) JPMARC マニュアルによるマッピング: サブフィールド単位で関係する NCR 条項番号がマニュアル上に記されていることについては前述した。これらの記載に基づき、JPMARC データ要素から NCR エレメントへのマッピングを導出した。JPMARC データ要素 245¥a, n, p を NCR エレメント「本タイトル」に対応づけることを第1表では示している。ただし、本研究におけるマッピングに揃えるため、サブフィールド n と p は限定子がある行に対応づけている。JPMARC ではインディケータを組み合わせたマッピングを示しておらず、そのため略記する場合にもインディケータを含めていない。これらゆえ、マッピングテーブル内で「JPMARC マニュアル」の欄に記載がない、あるいは記載があってもフィールド識別子(第3列)とサブフィールド識別子(第6列)の組み合わせと異なる指定内容の場合には、本研究によるマッピングと相違することになる。

加えて、第1表には該当する部分が出現していないが、JPMARC マニュアルが指示するマッピングが本研究のそれと大きく相違する場合がある。そのときには、列「JPMARC データ要素名」と「JPMARC 独自マッピング: NCR エレメント」にそのマッピング先を記載した。第1表には該当する行が出現しないため、これらの列自体を削除している(本稿末尾に記した URL から取得できるマッピングテーブルには、これらの列が含まれている)。

具体例を挙げると、MARC データ要素 300¥##¥e (キャリアに関する情報。付属資料)を、本研究では限定子「付属資料」を付けて「#2.41 体現形に関する注記」に対応づけた。それに対して、JPMARC マニュアルでは、上記データ要素を「#43.1 著作間の関連」と「#43.3 体現形間の関

連」に対応づけており、本研究とは見解の相違がある。つまり、付属資料を本体から独立した著作および体现形と捉えた上で関連として記録することを、JPMARC マニュアルは指示している。一方、本研究では、上記データ要素により記録される付属資料をそれ自体で独立したものと捉え記述することは困難と考え、注記に割り当てることにした。同様の両者の相違は相当数にのぼり、マッピングテーブルに単純に挿入した行数を数えると26であった。

2) RIMMF3によるマッピング: RDA エlementに対して示された MARC21 からのマッピングを、対応する NCR エlement に対するマッピングと見なした上で、列「RIMMF3: MARC21 から RDA」に簡略に記載した。RDA と等価な NCR エlement がテーブルに出現している範囲で、RIMMF3 が採用している「フィールド識別子+第1・第2 インディケータ+サブフィールド識別子」の組み合わせをそのまま転記した。それゆえ、JPMARC では採用されていないフィールドやサブフィールドが記載されている場合がある。なお、NCR エlement に限定子を付け行を分けて記載してある場合には、意味的に対応する行にのみ RIMMF3 によるマッピングを記載した。

245\$.Ya, 245\$.Yn, 245\$.Yp は、それぞれ「#2.1.1 本タイトル」、その限定子「巻次等」、「部編等のタイトル」にマッピングしている。これらは本研究におけるマッピングと等しい。一方、「#2.1.2 並列タイトル」に相当する RDA エlement parallel title proper は、290\$.1Ya からのマッピングとされている。MARC21 フォーマットには 290 というフィールドはなく、他にも RIMMF3 によるマッピングでは、292 など MARC21 に登録されていないフィールドが見受けられるが、その理由は不明である。

また、本研究で策定したテーブルには出現していない RDA エlement (すなわち現行 JPMARC には出現しない NCR エlement に相当する RDA エlement) に対応づけられている、多数の MARC21 データ要素は記載していない。

かつ、固定長フィールドからのマッピングについても、テーブル内には記載していない。

### III. 策定したマッピングの概要と課題整理

本章では、由来情報の記載、そして本研究で策定したマッピングの特徴である「読み」と「情報源」の扱いについて論じる。その後、JPMARC データ要素から NCR エlement へのマッピングにおけるカーディナリティに基づき、1対1, 1対多, 多対1に分けて、マッピングの該当例、該当ペア数、および課題や対処策等を整理する。ここでのマッピングの単位は前述した通り、MARC データ要素は可変長フィールドの場合、フィールド識別子からサブフィールド識別子までを組み合わせたものとなり、NCR2018 側は限定子を除いたエlement単位とし、これらの間のカーディナリティに基づく整理とする。マッピング後の調整処理についても併せて説明する。

また、説明用に1つの JPMARC レコード例を第1図に示し、その変換結果の表示例を第2図と第3図に示す。変換結果の第2図では、体现形、個別資料など、実体ごとに区切った表示としており、それら実体の下では各行の左から NCR エlement、限定子、値の順にタブ記号で区切り表示している。限定子がない場合には当該箇所は空値としてある。また、第3図は限定子と値の末尾の由来情報を削除するなどした表示形式である。紙幅の制約のため、体现形の記述部分のみ表示した。他の JPMARC レコードとその変換例8万件を本稿末尾に記した URL にて公開している。多様な事例については、それらを参照されたい。

#### A. 由来情報の記載、「読み」と「情報源」のマッピング

JPMARC のいかなるデータ要素からマッピングされた値であるのか判明するよう、元の値の出現箇所を記録したものを、本研究では「由来」(provenance) と呼ぶ。第2図の変換後のメタデータ表示において、エlement値の末尾に当該値の由来情報を“{245\$00Ya}”のように“{.}”で囲み表示している。これによって、例えば

JAPAN/MARC 書誌レコードから NCR2018 エレメントベースのメタデータへの変換

000	01409nam a2200409zi 4500	650 7  6 880-03  a 図学  0 00574954  2
001	032071450	ndlsh
003	JTNDL	7001  6 880-04  a 平野, 元久  e 著者
005	20220615145759.0	0 032197708
007	ta	7001  6 880-05  a 吉田, 一朗  e 著者
008	220422s2022 ja    g	0 032197719
	jpn	88000  6 245-01/\$1  a ワカル ズケイ カガ
015	a 23690253  2 jnb	ク
020	a 978-4-339-04677-9 :  c 2500 円	88000  6 245-01/(B  a Wakaru zukei
035	a (JP-ToTOH)34328004	kagaku
040	a JTNDL  b jpn  c JTNDL  e	880 1  6 264-02/\$1  b コロナシャ
	ncr/2018	880 1  6 264-02/(B  b Koronasha
084	a MA93  2 kktb	880 7  6 650-03/\$1  a ズガク  0 00574954
084	a 414.6  2 njb/10	2 ndlsh
090	a MA93-M4	880 7  6 650-03/(B  a Zugaku  0 00574954
24500	6 880-01  a わかる図形科学 /  c 平	2 ndlsh
	野元久, 吉田一朗 共著	8801  6 700-04/\$1  a ヒラノ, モトヒサ
264 1	6 880-02  a 東京 :  b コロナ社,  c	0 032197708
	2022.5	8801  6 700-04/(B  a Hirano, Motohisa
300	a 178p :  c 21cm	0 032197708
336	a テキスト  2 ncrcontent	8801  6 700-05/\$1  a ヨシダ, イチロウ
337	a 機器不用  2 ncrmedia	0 032197719
338	a 冊子  2 nrcarrier	8801  6 700-05/(B  a Yoshida, Ichiro  0
504	a 文献あり 索引あり	032197719

第1図 MARC レコード例

「#02.01.01 本タイトル」の値“わかる図形科学”が245\$00\$aからマッピングされた値であることが示されている。こうした由来情報により、適用されたマッピングが容易に理解できる（マッピングテーブルをその都度参照する必要がない）、記録されたエレメント値の意味も容易に把握でき、場合によっては不適切なマッピングあるいは不十分な変換処理に気づくことにもなる。むしろ、由来情報を表示しないという選択肢もあり、第3図による表示形式では表示していない。

記述フィールドの値に対する「読み」は、JPMARCではフィールド880（代替文字種表現）にまとめて記録され、それぞれ対象とする値との対応関係がサブフィールド6の下に記されている（第1図）。同時に、「読み」付与の対象となる個々のフィールド内でもサブフィールド6を用いて、880フィールドへの対応関係が示されている。NCR2018において、こうした値の「読み」は独立したエレメントには設定されていないが、表記とその読みが分離している日本語のデータにおいて、これまでの考え方を踏襲すればデータ要素としての必要性がある。さらには、著作のタイトルや個人・団体等を含めた典拠形APなどを構

成する際にも読みが不可欠な要素となる。

そこで本研究では、「読み」の付与対象の値が記録されているMARCフィールド内に擬似的に「読み」を収めるサブフィールドを設け、そこに読みを移行した後に、NCRエレメントへのマッピングに加えるという方策を採用した。具体的には、フィールド245のサブフィールドa, b, c, n, pに対して、それぞれの「読み」を収めるサブフィールドA, B, C, N, P（対応するサブフィールドの識別子を大文字表記したもの）を仮に設け、そこに読みを格納した。同じように、フィールド700（典拠形AP（創作者等）個人）、710（同。団体）、720（非統制形AP。創作者等）、730（著作）、740（非統制形AP。タイトル）、800（典拠形AP。シリーズの創作者等。個人）、810（同。団体）における「読み」についても仮想的にサブフィールドを設け、そこに格納した。

このように移行した上で、「読み」についてもNCRエレメントへのマッピングに含め、対応するエレメントにそのまま対応づけている。ただし、「読み」であることが判明するよう限定子「読み」、「巻次等読み」や、「個人名読み」、「世系読み」などを付けた。第1表のマッピングテー

# 表現形

#02.01.01 本タイトル わかる図形科学 {245¥00¥a}

#02.01.01 本タイトル 読み ワカル ズケイ カガク {245¥00¥A}

#02.02.01 本タイトルに関する責任表示 平野元久, 吉田一朗 共著 {245¥00¥c}

#02.05.01 出版地 東京 {264¥#1¥a}

#02.05.01 出版地 出版国コード ja {008/15}

#02.05.03 出版者 コロナ社 {264¥#1¥b}

#02.05.05 出版日付 2022.5 {264¥#1¥c}

#02.05.05 出版日付 西暦年 2022 {008/07}

#02.12 刊行方式 単巻資料 {000/07}

#02.15 機器種別 (ncrmedia) 機器不用 {337¥##¥a}

#02.16 キャリア種別 (ncrcarrier) 冊子 {338¥##¥a}

#02.17 数量 178p {300¥##¥a}

#02.18 大きさ 21cm {300¥##¥c}

#02.34 表現形の識別子 ISBN ISBN 978-4-339-04677-9 {020¥##¥a}

#02.34 表現形の識別子 他 MARC 番号等 他 MARC 番号等 (JP-ToTOH)34328004 {035¥##¥a}

#02.34 表現形の識別子 全国書誌番号 全国書誌番号 23690253 {015¥##¥a}

#02.35 入手条件 2500 円 {020¥##¥c}

#42.07 表現形から個別資料への関連 NDL 請求記号 MA93-M4 {090¥##¥a}

# 個別資料

#03.05 個別資料の識別子 NDL 請求記号 NDL 請求記号 MA93-M4 {090¥##¥a}

# 著作

#04.01 著作の優先タイトル (仮) わかる図形科学 || ワカル ズケイ カガク {245¥00¥a}{245¥00¥A}

#04.04 著作の日付 2022 {008/07}

#04.21 対象利用者 対象利用者コード 一般 {008/22}

#22.01 著作に対する典拠形アクセス・ポイント (仮) 平野, 元久 || ヒラノ, モトヒサ ; 吉田, 一朗 || ヨシダ, イチロウ . わかる図形科学 || ワカル ズケイ カガク {700¥1#}{700¥1#}{245¥00¥a}{245¥00¥A}

#44.01.01 創作者 著者: 平野, 元久 || ヒラノ, モトヒサ (NDL 典拠 ID 032197708) {700¥1#}

#44.01.01 創作者 著者: 吉田, 一朗 || ヨシダ, イチロウ (NDL 典拠 ID 032197719) {700¥1#}

#45 資料と主題との関連 件名. 普通件名 (ndlsh) 図学 {650¥#7}

#45 資料と主題との関連 分類記号 (kktb) MA93 {084¥##¥a}

#45 資料と主題との関連 分類記号 (njb/10) 414.6 {084¥##¥a}

# 表現形

#05.01 表現種別 (ncrcontent) テキスト {336¥##¥a}

#05.02 表現形の日付 2022 {008/07}

#05.03 表現形の言語 言語コード jpn {008/35}

#05.16 付加的内容 書誌注記 文献あり 索引あり {504¥##¥a}

#23.01 表現形に対する典拠形アクセス・ポイント (仮) 平野, 元久 || ヒラノ, モトヒサ ; 吉田, 一朗 || ヨシダ, イチロウ . わかる図形科学 || ワカル ズケイ カガク . テキスト . 2022 . jpn {700¥1#}{700¥1#}{245¥00¥a}{245¥00¥A}{336¥##¥a}{008/07}{008/35}

# 個人

#06.01 個人の優先名称 平野, 元久 {700¥1#¥a}

#06.01 個人の優先名称 読み ヒラノ, モトヒサ {700¥1#¥A}

#06.18 個人の識別子 典拠 ID 032197708 {700¥1#¥0}

#26.01 個人に対する典拠形アクセス・ポイント 平野, 元久 || ヒラノ, モトヒサ {700¥1#}

JAPAN/MARC 書誌レコードから NCR2018 エレメントベースのメタデータへの変換

```
# 個人
#06.01 個人の優先名称      吉田, 一朗 {700¥1#¥a}
#06.01 個人の優先名称 読み  ヨシダ, イチロウ {700¥1#¥A}
#06.18 個人の識別子      典拠 ID  032197719 {700¥1#¥0}
#26.01 個人に対する典拠形アクセス・ポイント      吉田, 一朗 || ヨシダ, イチ
        ロウ {700¥1#}

# データ管理情報
#レコード作成機関      JTNDL {040¥##¥a}
#レコード変換機関      JTNDL {040¥##¥c}
#レコード更新日付      20220615145759.0 {005}
#レコード管理番号      032071450 {001}
#全国書誌作成機関      jnb {015¥##¥2}
#目録用語      jpn {040¥##¥b}
#目録規則      ncr/2018 {040¥##¥e}
#資料区分      文字資料 {000/06}
#資料区分      文字資料 {007/00}
#資料区分(下位)      普通活字 {007/01}
```

第2図 MARC レコード例の変換結果. 続き

```
# 体現形
#02.01.01 本タイトル      わかる図形科学
#02.02.01 本タイトルに関する責任表示      平野元久, 吉田一朗 共著
#02.05.01 出版地      東京
#02.05.01 出版地      ja
#02.05.03 出版者      コロナ社
#02.05.05 出版日付      2022.5
#02.05.05 出版日付      2022
#02.12 刊行方式 単巻資料
#02.15 機器種別 機器不用
#02.16 キャリア種別      冊子
#02.17 数量      178p
#02.18 大きさ      21cm
#02.34 体現形の識別子      ISBN 978-4-339-04677-9
#02.34 体現形の識別子      他 MARC 番号等 (JP-ToTOH)34328004
#02.34 体現形の識別子      全国書誌番号 23690253
#02.35 入手条件 2500 円
#42.07 体現形から個別資料への関連      NDL 請求記号 MA93-M4
```

第3図 MARC レコード例の変換結果 (部分): 表示方法の変更

ブル上では、245¥00¥A や 245¥00¥N から「#2.1.1 本タイトル」の限定子「読み」や「巻次等読み」へのマッピングとしてある。これらの結果、第1図のフィールド880, サブフィールド6の値“245-01/\$1”に続く、サブフィールドaの値である“ワカル ズケイ カガク”が、第2図では「#02.01.01 本タイトル」の限定子「読み」の値として表示されている。

次に、「情報源」の扱いについて説明する。フィールドによっては、サブフィールド2「情報源」をもつ。これは当該フィールド内の他のサブフィールドが取り得る値が統制語彙などによって規定されている場合に、その用語群の名称

(通常は略称)を示す。RDAにおけるVES (Vocabulary Encoding Scheme; 語彙エンコーディングスキーム)と同じと捉えることができる。フィールド336 (表現種別), 337 (機器種別), 338 (キャリア種別), 340 (媒体の特性等), 344~347 (キャリアに関する情報), さらには600~651 (典拠形APのうち件名に該当するもの), およびジャンル・形式用語(655)において、こうしたサブフィールド2が出現する。336¥##¥aの値の情報源は「ncrcontent」に固定され、NCR2018が規定する表現種別の用語から選ばれた値であることを示している。同じく、337¥##¥aの値の情報源は「ncrmedia」、338¥##¥aの値の情報源は

「ncrcarrier」となる(第1表)。また、600~651の値の情報源は「ndlsh」(NDL 件名標目表)、655は「ndlgft」(NDL ジャンル・形式用語表)となる。

本研究における独自の扱いとして、NCR エレメントの限定子に「情報源」を設け、そこにこれらの値(情報源の略称)をマッピングした。すなわち、「#5.1 表現種別」や「#2.15 機器種別」、「#2.16 キャリア種別」に限定子「情報源」を付けてマッピングした(第1表)。同様に、600~651の場合には、NCR エレメント「#45 資料と主題との関連」の下で、限定子「件名. 個人. 情報源」、「件名. 団体. 情報源」などとして、通常の限定子に追記する形式で情報源を明示する限定子とした。分類記号の分類法についても情報源と同じ扱いとし、「#45 資料と主題との関連」の下の限定子「分類記号. 情報源」とした。

マッピング後の調整処理において、上記の限定子「情報源」を伴って記録された値(例: “#02.15 機器種別 情報源 ncrmedia {337¥##¥2}”)と、その対象となる値(例: “02.15 機器種別 機器不用 {337¥##¥al}”)とを組み合わせ合わせて合成し、単一の行の表示を構成している(例: “#02.15 機器種別 (ncrmedia) 機器不用 {337¥##¥al}”) (第2図)。つまり、「情報源」の対象となる値の限定子を「(ncrcontent)」, 「(ncrmedia)」など、あるいは「件名. 個人 (ndlsh)」, 「分類記号 (kktb)」などに置換した上で表示している。これら限定子を除去して表示することも可能であり、第3図に示した表示がそれに該当する。

Coyle による区分を適用すれば、これら「読み」と「情報源」とは、「あるサブフィールドは他のサブフィールドの値を説明あるいは補足するという連鎖の構成をとる場合」に該当する<sup>35)</sup>。

## B. 「1対1」のマッピング

マッピング元とマッピング先で意味的な設定粒度が等しくかつ1対1のマッピングとなる事例には、例えば JPMARC データ要素 000/07 (レコード種別) から「#2.12 刊行方式」へのマッピング、020¥##¥c から「#2.35 入手条件」へのマッ

ピングなどが挙げられる。

次に、マッピング元と先のデータ要素間で意味範囲が包含関係にあり、かつマッピング先であるエレメントがより範囲が広い事例には、511¥0#¥a (出演者注記) から「#2.41.2 責任表示に関する注記」へのマッピングなどが該当する。現行の JPMARC では他に同じエレメントに対応づけるデータ要素がないため、結果的に1対1となっている。なお、ここでは限定子「出演者注記」を付けてマッピングしているが、NCR エレメントに基づく表示の際には単純に限定子を除去して問題ない。

1対1となるマッピングペアは19にとどまり、優先度が低いとしたマッピングを除外した場合には26ペアに増加する(第2表)。つまり、優先度が低いペアをも含めたすべてのペアを対象にして集計した場合に、1対1以外の他のカーディナリティに位置づけられていたものが、優先度が低いペアを除外した結果、形成していたクラスタが分解され、1対1のペアが増えたことを意味する。

## C. 「1対多」のマッピング

意味上の設定粒度が等しくかつ1対多となる事例は例外的な位置づけである。008/07 (出版等の日付) から、「#2.5.5 出版日付」(限定子「西暦年」)に加えて、「#4.4 著作の日付」と「#5.2 表現形の日付」を生成する事例などが該当する。JPMARC データ要素には著作の日付、表現形の日付に直接該当する項目がなく、NCR2018 に依拠して著作と表現形の典拠形 AP を生成する際の識別情報としてこれら日付が望まれるため、便宜的に出版日付を流用した。

データ要素の意味範囲が包含関係、かつマッピング先がより狭い意味をもつ1対多の事例には、以下の2つの類型がある。

1) 元のデータ要素の意味を分解して、その該当するすべてにマッピングする場合。044¥##¥a (出版・製作国コード) を、「#2.5.1 出版地」、「#2.6.1 頒布地」、「#2.7.1 製作地」の3つのエレメントに限定子「MARC 国名コード」を付けてマッピングすることが当てはまる。同様に、550¥##¥a

(出版表示等に関する注記) の、「#2.41.5 出版表示に関する注記」と「#2.41.6 頒布表示に関する注記」へのマッピングも当てはまる。これらは、マッピング後の調整処理の段階において、出版・頒布・製作のいずれが妥当な事例であるのかを、他のエレメントである出版地、頒布地や出版者等の値の出現と照合して判定し、不用なマッピング結果の値を削除している。

また、245¥00¥b (並列タイトル, タイトル関連情報等) も同様な事例と見ることができる。このデータ要素の値は、並列タイトルの場合とタイトル関連情報の場合とがある。しかし、並列タイトルは 246¥[023]1¥a (第 1 インディケータは 0, 2, 3 のいずれかという指定) にも記録されるため、マッピング後の処理において、246 からマッピングされた値と照合することにより、上記の値が並列タイトルとタイトル関連情報のいずれに該当するのかを判定し、重複した値は削除している。

2) 優先度を明示して、複数エレメントにマッピングする場合。データ要素 245¥00¥c や 250¥##¥a, 250¥##¥b, あるいは 500¥##¥a などにおける優先度を設定したマッピングについては、前述した通りである。マッピング後に、それらすべての値を採用し表示するのか、あるいは指定された場合のみ優先度の低いマッピングの結果を含めて表示するのかなどが、事後の調整処理となる。第 2 図に示した表示例では、優先度が低いマッピング結果は含めていない。仮に優先度が低いものを含めたときには、「#02.05.02 並列出版地\* 東京 {264¥#1¥a}」や「#02.05.04 並列出版者\* コロナ社 {264¥#1¥b}」などが加わることになり、結果的にノイズである場合が多い。

1 対多のマッピングペア数は 298 となった (第 2 表)。優先度が低いとしたマッピングを除外したときには、166 ペアに大きく減少した。特にエレメント限定子をもたないマッピングペアが 162 から 38 へと大幅な減少を見せている。

#### D. 「多対 1」のマッピング

データ要素単位の設定粒度が等しくかつ多対 1 のマッピングとなる事例には、リテラル値をとる

JPMARC データ要素とコード値をとるデータ要素などが同じマッピング先エレメントをもつ場合が該当した。264¥[#23]1¥a (出版地等), 008/15 (出版国コード), 044¥##¥a (出版・製作国コード) はいずれも「#2.5.1 出版地」にマッピングされる (第 2 図)。ただし、後者の 2 つはコード値であることを限定子によって示す。同様に、008/18 と 310¥##¥a とは、いずれも「#2.13 刊行頻度」にマッピングされる。

マッピング先のデータ要素の意味範囲がより広い事例は多数存在する。これらには限定子を適用し、マッピング後の調整処理において限定子を参照するという方策を採用した。

1) 245¥00¥a に加えて、245¥00¥n (巻次等), 245¥00¥p (部編等のタイトル), 245¥00¥t (2 番目以降の個別タイトル等) は、すべて「#2.1.1 本タイトル」にマッピングされる。これらを相互に区別し、その後の最終的な表示形を構成するなどを用意して、245¥00¥a 以外は限定子を付けてマッピングしている。「#2.34 体現形の識別子」をマッピング先とする多数のデータ要素についても同様である。020¥##¥a (ISBN), 020¥##¥z (無効な ISBN), 022¥[#0]¥a (ISSN), その他多数のデータ要素をそれぞれの限定子を付けて上記エレメントにマッピングした。Coyle による区分を適用すれば、これは「個々のサブフィールドが相互に独立して対象リソースを記述する場合」に該当する<sup>35)</sup>。

2) 実体間の関連を記録する「#43.1 著作間の関連」, 「#43.2 表現形間の関連」, 「#43.3 体現形間の関連」も多数の JPMARC データ要素からのマッピングとなる。例えば、「#43.1 著作間の関連」を見ても、510¥2#¥a (索引サービスの名称), 760¥0#¥t (上位のシリーズ, リンク先レコードのタイトル), 760¥0#¥w (同, リンク先レコードの管理番号), 760¥0#¥x (同, 上位のシリーズの ISSN), 762¥0#¥t (サブシリーズ, リンク先レコードのタイトル) などからマッピングされる。加えて、780¥0[014567]¥t (先行, リンク先レコードのタイトル), 785¥0[014567]¥t (後続, リンク先レコードのタイトル) など、多数のデータ要素

からのマッピングとなる。ただし、これらに対する限定子は MARC サブフィールドに依拠したものではなく、NCR2018 による関連指示子を採用し、場合によってはさらに付加情報を加えた。上記の事例の場合、それぞれの限定子は「索引サービス (著作)」、「上位のシリーズ」、「上位のシリーズ. 書誌 ID」、「上位のシリーズ. ISSN」、「サブシリーズ」、「先行 (著作)」、「後続 (著作)」とした。マッピング後の調整処理において、これらに関連指示子として値に前置する形式に限定子から移行させている。もちろん、限定子は MARC サブフィールドをそのまま流用し、その後の段階で関連指示子等に置換するという方策も取り得る。

これに関して付言すると、データ要素 780  $\$0[014567]\$g$  (改題発生巻次・年月次) などは、厳密に捉えるならば、関連の値として記録すべき範囲をはみ出しているといえよう。しかし、本研究では記録に含めるべき事項を広く捉え、マッピングし表示することとした。これらはマッピングから除外するという選択肢もあり、マッピング策定における相違が発生しうる。

3) 当該資料の主題を表す分類記号や件名標目などは、すべて「#45 資料と主題との関連」へのマッピングとした (第 2 図)。NCR2108 では当該章は「保留」とされ一切の規定がないが、本研究ではこれを単一の要素と見立てた上で、著作に付随する要素に位置づける。すべてのデータ要素を「#45 資料と主題との関連」にマッピングした後に、組み立てて元の件名標目等を形成した。

多対 1 となるマッピングペアは、総計 687 となった (第 2 表)。優先度が低いペアを除外した場合にも 649 ペアであり、減少幅は小さい。

### E. 「多対多」のマッピング

多対多のパターンに該当するマッピングペアは、第 2 表の集計では 1 対多と多対 1 に含めて数えられており、多対多のパターンと重複した数え方となっている。そこで、1 対多と多対 1 の合計ペア数から多対多のペア数を引き、それに 1 対 1 のペア数を加えると、総数に一致することになる

(298+687-217+19=787)。

多対多のマッピングとは、1) 結果的に形成される場合と、2) 構造的な特徴により、いわば意図されて形成される場合とがある。

1) 例えば 245 $\$00\$\$b$  は「#2.1.2 並列タイトル」と「#2.1.3 タイトル関連情報」とにマッピングされ (1 対多)、他方 246 $\$[023]1\$\$a$ , 246 $\$[023]1\$\$n$ , 246 $\$[023]1\$\$p$  からも「#2.1.2 並列タイトル」はマッピングされている (多対 1)。この「#2.1.2 並列タイトル」が結節点となり、全体として見たときには多対多の関係となる。他の事例では、044 $\$##\$\$a$  は「#2.5.1 出版地」、「#2.6.1 頒布地」、「#2.7.1 製作地」にマッピングする (1 対多) と同時に、264 $\$[##23]1\$\$a$  と 008/15 から「#2.5.1 出版地」にはマッピングがある (多対 1)。これらも、全体として見たときには多対多の関係となる。これらに対するマッピング後の調整処理については、前述した通りである。

2) 構造的な特徴に起因する多対多関連とは、次のものを指す。個人の典拠形 AP を構成する 700 $\$1[##2]\$\$a$  (個人. 優先名称) とそれに伴う生年・没年等のデータ要素は、「#6.1 個人の優先名称」と「#6.3 個人と結びつく日付」等の個人属性エレメントにマッピングするとともに、「#44.1.1 創作者」と「#44.2.1 寄与者」(限定子はいずれも「個人名」など) にもマッピングする (1 対多)。同様に、団体の典拠形 AP を構成する 710 $\$2[##2]\$\$a$  (団体. 優先名称) とその他データ要素は、「#8.1 団体の優先名称」等の団体属性エレメントにマッピングするとともに、「#44.1.1 創作者」と「#44.2.1 寄与者」(限定子はいずれも「団体名」など) にもマッピングする (1 対多)。これらのマッピングを逆に「#44.1.1 創作者」や「#44.2.1 寄与者」において集約すると、JPMARC データ要素からの多対 1 関係となる。これらゆえ、全体としては多対多のマッピングとなる。ここでは個人と団体という区分と創作者・寄与者という役割の区分とが交差しており、多対多の対応関係を形成している。なお、創作者と寄与者のいずれが事例に即して妥当であるのかは、記録されている「関連指示子」(700 $\$1[##2]\$\$e$ , 710 $\$1[##2]\$\$e$  な

ど)の値の参照によってマッピング処理後に判定する。

#### F. マッピング困難なデータ要素

マッピング困難なデータ要素には2種類ある。1つは、いずれの実体に対応するのか判定できないなど、NCR2018の構造に合致しないデータ要素である。こうしたデータ要素は、旧来のMARC21フォーマットを継承して使用していることによるものといえよう。NCR2018に従いメタデータを作成するとの観点からは、望ましくないデータ要素と位置づけられる。データ要素041¥[01]#¥h(言語コード, 原文の言語), 510¥2#¥b(索引サービス記事採録期間), 536¥##¥a(科研費情報, 補助金情報), 536¥##¥f(科研費情報, 科研費課題番号), 580¥##¥a(リンクに関する注記)などがこれに該当する。いずれの実体に割り当てべきか、さらにはその実体の下のいずれのエレメントが最も関係するのか不明であり、マッピング困難とした。例えば、科研費情報は体現形に伴うものか、あるいは著作に付随するものが判然としないため、マッピング困難とした。これらは見出し「# その他: 位置づけ不明なデータ要素」の下に表示する(第2図には該当例なし)。こうしたマッピング困難なデータ要素については、その必要性を含め、今後の再検討が求められる。

もう1つのマッピング困難なデータ要素群は、メタデータの管理情報に相当するものである。015¥##¥2(全国書誌作成機関), 040¥##¥a(レコード作成機関), 040¥##¥b(目録用言語), 040¥##¥c(レコード変換機関), 040¥##¥e(目録規則)などが該当する。これらは変換後の表示においては、見出し「# データ管理情報」の下に表示している(第2図)。

加えて、007/00と000/06に示された資料区分(コード値)や、007/01や008/21, 008/23などにおけるコード値も、それぞれ値ごとにマッピングすべき実体とエレメントが変わりうる。例えば、「映像・静止画資料」の場合、008/33は「資料内容種別」とされているが、そのコードは「静止画, 設計図, カード, スライド, 映像資料, コー

ド不使用」に分けられており、こうした区分は単一のNCRエレメントには収まらない。あるいは、個々のコード値が表すものが、NCR2018が規定する用語に合致しない場合も多い。これらはNCR2018適用以前からの「資料区分」とその下での各種コードが引き継がれて現在も残されているものと推測するが、NCR2018の規定には合致せず、再考が必要と考える。第2図では、これらを「# 資料区分」と「# 資料区分(下位)」として、便宜的に管理情報に位置づけた。

#### IV. 策定したマッピングに基づく JAPAN/MARC レコードの変換処理

策定したマッピングを参照して実際にJP-MARC 書誌レコードを変換するツールを試作した。ツールの構成と動作手順は、以下の通りである。

- 1) マッピングテーブルを読み込み、内部形式に変換する。これには、コード化フィールド用の変換テーブルも含まれる。
- 2) JPMARC 書誌レコードを読み込み、データ要素に分解した上で、内部形式に変換する。この段階で、「読み」の対象フィールドへの移行も行う。
- 3) マッピングテーブルと個々のMARCデータ要素とを照合し、NCRエレメントおよび限定子(ないときには空値)とその値に変換する。その際に、値の由来についても記録する。適切にマッピングできない要素も、それと分かる形式で保持する。
- 4) マッピング後の必要な調整処理を、エレメントごとに加える。典拠形APなど、複数要素を組み合わせて値を形成する処理なども、ここに含まれる。
- 5) NCR2018に基づく実体とエレメントによる形式に変換したメタデータを出力する。表示形式は複数設定可能であり、第2図や第3図による形式以外にもあり得る。

以下、マッピング後の調整処理など、開発したツールについて補足する。

- 1) 典拠形APなどを含めて、複数要素により

値を形成する処理は、基本的にエレメントごとに処理内容が異なる。「#4.1 著作のタイトル」は、フィールド 245 のサブフィールド a, b, n, p およびそれらの読みから機械的に生成している。そのため、限定子には「(仮)」を設定した(第2図)。「#22.1 著作に対する典拠形 AP」も、「#44.1.1 創作者」と上記の著作のタイトルを結合させて仮に生成している。他方、「#23.1 表現形に対する典拠形 AP」は、著作の典拠形 AP にさらに「#5.1 表現種別」、「#5.2 表現形の日付」、および「#5.3 表現形の言語」(限定子「言語コード」)の値を結合させて形成した。

2) 「#42 資料に関する基本的関連」に該当する「#42.6 体現形から表現形への関連」、「#42.1 著作から表現形への関連」などは、生成していない。学習用データとして、今回生成する必要性は高くないと判断したためである。ただし、フィールド 730 によって著作の典拠形 AP が記されているものは、それを採用し記録した。

併せて、開発したツールを用いて、一定量の JPMARC レコードの変換を試行した。NDL が週次更新により提供している JPMARC データのうち、2022 年 1 月から 7 月末までの新規または更新レコードを利用した<sup>37)</sup>。提供データから NCR2018 を用いて作成されたものに限定して抽出し、変換処理を行った。変換対象レコード数は 87,380 件であった。変換結果は第 2 図と第 3 図の表示形式それぞれによるものとし、別途 URL から公開している。

変換結果のエレメントごとの出現回数、由来ごとの出現回数を集計したところ、体現形が 87,380 件、「#2.1.1 本タイトル」87,553 件、「#2.1.2 並列タイトル」6,199 件、「#2.1.3 タイトル関連情報」42,550 件、「#2.1.4 並列タイトル関連情報」7 件となった。さらに、「#2.1.1 本タイトル」の由来ごとの集計では、245¥00¥a からマッピングされた値が 56,646 件、245¥00¥a と 245¥00¥n の組み合わせが 30,676 件、245¥00¥a と 245¥00¥p の組み合わせが 58 件、245¥00¥t が 173 件であった。体現形の総数と「#2.1.1 本タイトル」の出現回数の差分は、245¥00¥t (2 番目以降の個別のタイトル等)

に合致する。こうした集計によって、エレメントごとの値出現の全体的な傾向やその由来の細部を把握することができる。

## V. おわりに

本研究の目的は、NDL により NCR2018 を適用し作成された JPMARC 書誌レコードに対して、NCR2018 の構成に沿ったメタデータ、すなわち実体とそのエレメントからなるメタデータに再構成して表現すること、そのために必要なスキーマレベルのマッピングの策定とそれを活用したツールを開発し実行することである。

本研究では、先行事例を参照しつつも、独自にマッピングを策定した。マッピングは JPMARC レコードを変換し再構成することに目的を限定し、かつ JPMARC フォーマットで表現されている要素は基本的にすべて採用しマッピング対象とした。こうした方針の下、マッピングにおいて遭遇する個々の課題に対して適切な対処策を導入しそれらの解決を図った。具体的には、1) マッピングとその後の調整処理を切り分ける、2) マッピングにおいて NCR エレメントに適宜、限定子を導入し、その後の調整処理において活用する、3) JPMARC データ要素単位は「フィールド識別子、インディケータ、サブフィールド識別子」の組み合わせとし、場合によってはさらに「資料区分」を組み合わせ、より精度の高いマッピングとする、4) 優先度を設けて 2 レベルに分けたマッピングとする、5) 個々の値の由来情報を併せて記録するなどであり、これらによって課題群に対処した。

併せて、策定したマッピングについて、カーディナリティのパターンに分け、詳細にその該当例の検討や残された課題の整理などを行った。同時に、適切なマッピングが困難な部分の同定を図った。マッピングにおけるこうした詳細な検討は他に報告例が殆どなく、その点でも有効な成果と考える。これらは他のスキーマ間のメタデータマッピング、特に比較的詳細なメタデータ間でマッピングを検討する際に参考になるものと考え

さらに、策定したマッピングに基づいてデータ変換を実行するツールを実装し、一定量のデータ変換を試行した。これによって、広義の学習用コンテンツとしての NCR2018 によるデータ事例を大幅に増やすことができた。

他方、本研究はあくまでも NDL による NCR2018 適用細則および JPMARC フォーマットが採用しているデータ要素の範囲に限定される。フルスペックの MARC21 フォーマットと最大限の NCR2018 エレメントとの間でマッピングを検討したわけではない。また、変換結果のデータ群の学習用コンテンツとしての有効性などは確認できていない。そうした評価はかなり困難な課題と認識しており、今後の課題である。

上記課題の解決以外にも、今後の展開の方向性は複数考えられる。

1) NDL の JPMARC 典拠レコードに対して同様にマッピングテーブルの策定と変換ツールの実装を試みる、あるいはそれに代えて典拠データの提供サービス Web NDL Authorities に API でアクセスして必要な典拠データを取得し、本研究による書誌レコードからの変換結果と組み合わせるといふ方策もある。これによって、より内容豊かなデータ事例の提供を目指す。

2) 他の書誌レコードから抽出された著作や表現形の事例（インスタンス）との同一性の判定を実現する。加えて、単一 MARC レコードからの複数の著作や表現形への分割と抽出や、表現形間、著作間、表現形間の関連における関連先インスタンスの照合と同定なども課題である。

3) 変換後の NCR2018 エレメントベースのデータから、RDF 形式のデータを生成する。その際には、a) 基本的に JLA 目録委員会が公開している RDF データ例<sup>38)</sup> に沿ったものとする、b) 実体インスタンスごとに仮の URI を生成し、それを用いて実体間のリンクを形成する、c) 値語彙が既に NCR2018 に登録されているものは、値の表現を URI で置換する、d) マッピング困難としたデータ要素は除外し RDF データには含めないなどの扱いとする。

本研究で策定したマッピングテーブル、および

JPMARC レコードとその変換結果のデータセットを下記 URL から公開している。適宜、参照されたい。

<https://github.com/ShoichiTaniguchi/NCR2018-LOD>

## 謝 辞

筆者は現在、JLA 目録委員会の委員を務めており、本研究も委員会業務から着想を得た研究課題であるが、あくまでも個人研究であり、委員会としての見解を示したものではない。こうした研究の契機を与えてくれたこと、および成果発表を寛容に認めていただいていることに対して、目録委員会の委員諸氏に謝意を表す。

## 注・引用文献

- 1) 日本図書館協会目録委員会. 日本目録規則. 2018 年版, 日本図書館協会, 2018, 761p. <https://www.jla.or.jp/mokuroku/ncr2018>, (入手 2022-07-26).
- 2) 日本図書館協会目録委員会. 日本目録規則 2018 年版データ作成事例. 2021-, <https://www.jla.or.jp/mokuroku/ncr2018-sample>, (入手 2022-07-26).
- 3) 木村麻衣子編著. 『日本目録規則 2018 年版』入門. 日本図書館協会, 2022, 149p.
- 4) 鳥海恵司編. 日本目録規則 2018 年版体現形 (書誌) データ完全実例集: 音楽とその周辺. トッカータ, 2020, 158p.
- 5) 鳥海恵司編. 日本目録規則 2018 年版典拠データ完全実例集: 音楽とその周辺. トッカータ, 2020, 376p.
- 6) “資料 NCR2018 によるデータ作成事例”. 情報資源組織論. 三訂, 田窪直規編著. 樹村房, 2020, p. 203-216.
- 7) 国立国会図書館. 日本目録規則適用細則類一覧. 2021-. [https://www.ndl.go.jp/jp/data/catstandards/ncr\\_regulations/index.html](https://www.ndl.go.jp/jp/data/catstandards/ncr_regulations/index.html), (入手 2022-07-26).
- 8) 国立国会図書館. JAPAN/MARC MARC21 フォーマットマニュアル (2021 年 1 月). 単行・逐次刊行資料編, 2022 年 4 月 25 日更新. 国立国会図書館, 2022, 317 p. [https://www.ndl.go.jp/jp/data/JAPANMARC\\_MARC21manual\\_MS\\_202101.pdf](https://www.ndl.go.jp/jp/data/JAPANMARC_MARC21manual_MS_202101.pdf), (入手 2022-07-26).
- 9) RIMMF3 Home. <https://www.rimmf.com/w/doku.php?id=rimmf>, (入手 2022-07-26).
- 10) Dunsire, Gordon; Fritz, Deborah; Fritz, Richard. “Instructions, interfaces, and interoperable data: The RIMMF experience with RDA”. IFLA WLIC

- 2016, Columbus, Ohio, 2016, 8p. <http://library.ifa.org/id/eprint/1324/1/093-dunsire-en.pdf>, (入手 2022-07-26).
- 11) Dunsire, G.; Fritz, D.; Fritz, R. Instructions, interfaces, and interoperable data: The RIMMF experience with RDA revisited. *Cataloging & Classification Quarterly*. 2020, vol. 58, no. 1, p. 1-15.
  - 12) フォルダRIMMF3>tables内のファイル“en\_rda+tmp\_elements.txt”と“marc-fixedfields.txt”が該当する。
  - 13) RDA Steering Committee. MARC 21 Format for Bibliographic Data Mapped to RDA Elements. <https://original.rdatoolkit.org/document.php?id=jscmap2>, (入手 2022-07-26).
  - 14) RDA Registry. Map From RDA Properties to MARC 21 Bibliographic Encodings. <http://www.rdaregistry.info/Maps/mapRDA2M21B.html>, (入手 2022-07-26).
  - 15) RDA Registry. Alignment From RDA Elements to MARC 21 Bibliographic Encodings. <http://www.rdaregistry.info/Aligns/alignRDA2M21B.html>, (入手 2022-07-26).
  - 16) Yee, M. M. FRBRization: A method for turning online public finding lists into online public catalogs. *Information Technology and Libraries*. 2005, vol. 24, no. 3, p. 77-95.
  - 17) Manguinhas, Hugo; Freire, Nuno; Borbinha, José. “FRBRization of MARC records in multiple catalogs”. *JCDL'10: Proceedings of the 10th Annual Joint Conference on Digital Libraries*, 2010, p. 225-234. <https://doi.org/10.1145/1816123.1816157>, (入手 2022-07-26).
  - 18) Decourselle, Joffrey; Duchateau, Fabien; Lumineau, Nicolas. “A Survey of FRBRization techniques”. *Theory and Practice of Digital Libraries (TPDL)*, 2015, Poznań, Poland, p. 185-196. <https://hal.archives-ouvertes.fr/hal-01198487/file/SurveyFRBRization.pdf>, (入手 2022-07-26).
  - 19) RDA Steering Committee. Complete Examples-Bibliographic Records, April 2016. [https://www.rdatoolkit.org/sites/default/files/rsc\\_rda\\_complete\\_examples\\_bibliographic\\_april2016\\_0.pdf](https://www.rdatoolkit.org/sites/default/files/rsc_rda_complete_examples_bibliographic_april2016_0.pdf), (入手 2022-07-26).
  - 20) Maxwell, R. L. *Maxwell's Handbook for RDA*. ALA, 2013, 900p.
  - 21) 国立国会図書館. MARC21とDC-NDL (RDF)の書誌データの主な違い. 2016. [https://www.ndl.go.jp/jp/data/data\\_service/jnb/pdf/MARC21andDC-NDL.pdf](https://www.ndl.go.jp/jp/data/data_service/jnb/pdf/MARC21andDC-NDL.pdf), (入手 2022-07-26).
  - 22) Library of Congress. MARC 21 Formats. MARC Mappings. 2014-. <https://www.loc.gov/marc/marcdocz.html>, (入手 2022-07-26).
  - 23) ISO 25964-2: 2013. *Thesauri and Interoperability with Other Vocabularies, Part 2: Interoperability with Other Vocabularies*. ISO/TC46/SC9, 2013, 99p.
  - 24) Haslhofer, Bernhard. *A Comparative Study of Mapping Solutions for Enabling Metadata Interoperability*. TR-20080202, University of Vienna, 2008, 38p. <https://eprints.cs.univie.ac.at/295/1/covered.pdf>, (入手 2022-07-26).
  - 25) Haslhofer, B.; Klas, W. A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys*. 2010, vol. 42, no. 2, article 7, 37p.
  - 26) Woodley, Mary S.; Baca, Murtha. “Metadata matters: Connecting people and information”. *Introduction to Metadata*. 3rd ed. Getty Publications, 2016. <https://www.getty.edu/publications/intrometadata/metadata-matters/>, (入手 2022-07-26).
  - 27) Aichison, Jean; Gilchrist, Alan. シソーラス構築法. 内藤衛亮ほか訳. 丸善, 1989, p. 118-121.
  - 28) Chen, Y. N. A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level. *Library Hi Tech*. 2015, vol. 33, no. 2, p. 175-194.
  - 29) Halpin, Harry et al. “When owl:sameAs isn't the same: An analysis of identity in Linked Data”. *The Semantic Web—ISWC 2010. Lecture Notes in Computer Science*, vol. 6496. Springer, 2010, p. 305-320. [https://doi.org/10.1007/978-3-642-17746-0\\_20](https://doi.org/10.1007/978-3-642-17746-0_20), (入手 2022-07-26).
  - 30) SKOS Simple Knowledge Organization System Reference, W3C Recommendation, 18 August 2009. <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>, (入手 2022-07-26).
  - 31) Najjar, Jihad; Ternier, Stefaan; Duval, Erik. Interoperability of learning object repositories: Complications and guidelines. *IADIS International Journal on WWW/Internet*. 2004, vol. 2, issue 1. <http://www.iadisportal.org/digital-library/iadis-international-journal-on-www/internet>, (入手 2022-07-26).
  - 32) Chau, Do Van. *Challenges of Metadata Migration in Digital Repository: A Case Study of the Migration of DUO to Dspace at the University of Oslo Library*. Oslo University College, Master thesis, 2011. [https://oda.oslomet.no/oda-xmlui/bitstream/handle/10642/990/Chau\\_Do\\_Van.pdf](https://oda.oslomet.no/oda-xmlui/bitstream/handle/10642/990/Chau_Do_Van.pdf), (入手 2022-07-26).
  - 33) Library of Congress. MARC 21 Format for Bibliographic Data. 1999 Edition. Update No. 1 (October 2000) through Update No. 34 (July 2022).

JAPAN/MARC 書誌レコードから NCR2018 エレメントベースのメタデータへの変換

2022. <https://www.loc.gov/marc/bibliographic/>, (入手 2022-07-26).
- 34) Thomale, Jason. Interpreting MARC: Where's the bibliographic data? Code4lib Journal. 2010, issue 11. <https://journal.code4lib.org/articles/3832>, (入手 2022-07-26).
- 35) Coyle, Karen. MARC21 as data: A start. Code4lib Journal. 2011, issue 14. <https://journal.code4lib.org/articles/5468>, (入手 2022-07-26).
- 36) 007/00 による資料区分以外にも、フィールド 008 における 6 つの資料区分、さらにはレコードラベル 000 の位置 06 「レコード種別」における 9 区分がある。これらは意味上の重なりが見られるが、区分の仕方が微妙に異なり、かつ相互に異なるコードが採用されている。これらを使い分ける意図は理解できず、煩雑と言わざるをえない。この点に関する Coyle による同様の指摘もある。
- 37) 国立国会図書館. JAPAN/MARC データ (毎週更新). [https://www.ndl.go.jp/jp/data/data\\_service/jnb\\_product.html](https://www.ndl.go.jp/jp/data/data_service/jnb_product.html), (入手 2022-07-26).
- 38) 日本図書館協会目録委員会. NCR2018 年版エレメント・語彙等データ提供. RDF データの使用例. 2019-. <https://www.jla.or.jp/mokuroku/ncr2018-LOD>, (入手 2022-07-26).

## 要 旨

**【目的】** 国立国会図書館により日本目録規則 2018 年版 (NCR2018) を適用し作成された JAPAN/MARC (JPMARC) 書誌レコードに対して, NCR2018 の構成に沿ったメタデータすなわち実体とそのエレメントからなるメタデータに再構成すること, そのために必要なスキーマレベルのマッピングの策定とそれを活用したツールを開発し実行することを目的とする。これによって, 広義の学習用コンテンツとして利用可能な NCR2018 のデータ事例を大幅に増やすことができる。

**【方法】** 先行事例を参照しつつも, 独自にマッピングを策定した。マッピングは JPMARC データを変換し再構成することに目的を限定し, かつ JPMARC フォーマットで表現されている要素は基本的にすべてマッピング対象とした。こうした方針の下, マッピングにおける個々の課題に対して適切な対処策を導入しそれらの解決を図った。さらに, 策定したマッピングに基づいてデータ変換を実行するツールを実装し, 変換を試行した。

**【結果】** 策定したマッピングおよび変換用ツールは, 1) マッピングとその後の調整処理を切り分ける, 2) マッピングにおいて NCR エレメントに適宜, 限定子を導入し, その後の調整処理において活用する, 3) JPMARC データ要素単位は「フィールド識別子, インディケータ, サブフィールド識別子」の組み合わせとし, 場合によってはさらに「資料区分」を組み合わせ, より精度の高いマッピングとする, 4) 優先度を設けて 2 レベルに分けたマッピングとする, 5) 個々の値の由来情報を併せて記録することなどによって, 課題群に対処した。併せて, 策定したマッピングについて, カーディナリティのパターンに分け, 詳細にその該当例の検討や残された課題の整理などを行い, 適切なマッピングが困難な部分を同定した。また, 実装したツールによってデータ変換を実行し, 87,000 件のレコードを変換し, NCR2018 の構成に沿ったメタデータを作成することができた。