

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

谷口 祥一 (慶應義塾大学文学部)

taniguchi@z2.keio.jp

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 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を目的として、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の開発によって、87,380 件のレコードを変換した。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)優先度を設けて 2 レベルに分けたマッピングとするなど、各種の工夫を組み入れた。

1. はじめに

日本目録規則 2018 年版 (NCR2018) は、従来の 1987 年版と比べ内容や構成が大きく異なる。しかし、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。他方、国立国会図書館 (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。

本研究の目的は、NDL により NCR2018 を適用し作成された、JAPAN/MARC MARC21 フォーマット (JPMARC) によるレコードに対して、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を策定し、それを活用したツールを開発することである。これによって、メタデータ作成者が 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。すなわち、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。また、今回は JPMARC 書誌レコードのみ対象とする。

本研究の直接的な先行事例には、RDA の関

連ツール RIMMF3 がある。同ツールにより、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 の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。

2. マッピングの方針と策定

本研究で採用したマッピングの方針は、以下の通りである。

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」・「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。

3)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。併せて、NCR エレメントに適切にマッピングできないデータ要素も同定する。

4)マッピング自体とその後の調整処理を切り分け、必要な調整処理を検討し機能実装を図る。

5)逆方向のマッピングとなる NCR2018 エレメントに基づくメタデータを、MARC21 フォーマットに変換することは意図しない。

先行事例等を参照しつつ、本研究におけるマッピングのデータ単位を決定した。マッピング先となる NCR2018 のエレメントは最も特定性の高い下位レベルのエレメントとした。場合によっては、さらにエレメントを擬似的に細分する限定子を導入したが、この点は後述する。

JPMARC データ要素については、固定長フィールドと可変長フィールドに分かれ、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。また、サブフィールド内での区切り記号による分割も、部分的に採用した。

加えて、本研究で独自に、フィールド007の位置00(資料カテゴリー)の値による10区分を、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した。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のサブフィールドc(大きさ)を対応づけることができるが、#2.18.1と#2.18.2を除外するため、上記の資料区分「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。

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。テーブルには、左から順にNCRエレメント、限定子、それにマッピングするJPMARC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記載している。策定したマッピングテーブルを集計したところ、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。なお、この集計では資料区分の組み合わせは捨象している。

1)エレメント限定子の導入：NCRエレメントに比してJPMARC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。これら限定子の導入によって、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。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。

なお、マッピングテーブルには、参考までにJPMARCマニュアルによるマッピング、RIMMF3によるマッピングを併せて記載した。

3. 策定したマッピングの概要と課題整理

第1図に1つのJPMARCレコード例を、第2図にその変換結果(部分)を示した。

3.1 由来情報の記載、「読み」と「情報源」のマッピング

1)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、エレメント値の末尾に当該値の由来情報を“{245¥00¥a}”のように“{..}”で囲み表示している。これによって、例えば「#02.01.01 本タイトル」の値“わかる図形科学”が、フィールド245、第1・2インディケータ値0、サブフィールド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、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。むろん、由来情報を表示しないという選択肢も提供している。

2)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、対象とする値との対応がサブフィールド6に記されている。NCR2018において、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、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。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」などを付けた。3)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。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)など、さらには600~651

(典拠形アクセス・ポイントのうち件名に該当するもの)などにおいて、サブフィールド2が出現する。

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。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 {337Y##Y2}”)と、その対象となる値(例:“#02.15 機器種別 機器不用 {337Y##Ya}”)とを組み合わせることで合成し、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 {337Y##Ya}”)。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。むろん、これら限定子を除去して表示する選択肢も提供している。

3. 2 カーディナリティに基づくまとめ

1) 「1対1」のマッピング: マッピング元と先で意味的な設定粒度が等しくかつ1対1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「#02.12 刊行方式」へのマッピング、020Y##Yc から「#02.35 入手条件」へのマッピングなどが挙げられる。

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、511Y0#Ya (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。

2) 「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、「#04.04 著作の日付」と「#05.02 表現形の日付」を生成する事例などが該当する。

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、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は、マッピング後の調整処理の段階において、いずれが妥当な事例であるのかを、他のエレメント値の出現と照合して判定し、不要なマッピング結果の値を削除している。②優先度を明示して、複数エレメントにマッピング

する場合があるが、結果的にノイズであることが多い。

3) 「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。264Y[#23]1Ya (出版地等)、008/15 (出版国コード)、044Y##Ya (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。

他方、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。

4) 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。これらは、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 041Y[01]#Yh (言語コード. 原文の言語)、510Y2#Yb (索引サービス記事採録期間)、536Y##Ya (科研費情報. 補助金情報)、536Y##Yf (科研費情報. 科研費課題番号)、580Y##Ya (リンクに関する注記)などがこれに該当する。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。015Y##Y2 (全国書誌作成機関)、040Y##Ya (レコード作成機関)、040Y##Yb (目録用言語)、040Y##Yc (レコード変換機関)、040Y##Ye (目録規則)などが該当する。

4. JAPAN/MARC レコードの変換の試行

開発したツールを用いて、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、かつ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。対象レコード数は87,380件であった。変換結果は第2図の表示形式と、限定子や由来情報を除いた表示形式によるものとした。

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。

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

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

第1図 JPMARC レコード例

```

[NCR エレメント] [限定子] [マッピング値] {由来情報}
# 体現形
#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 (nrcarrier) 冊子 {338¥##¥a}
#02.17 数量 178p {300¥##¥a}
#02.18 大きさ 21cm {300¥##¥c}
#02.34 体現形の識別子 ISBN ISBN 978-4-339-04677-9 {020¥##¥a}
#02.34 体現形の識別子 他 MARC 番号等 他 MARC 番号等 (JP-ToTOH)34328004 {035¥##¥a}
#02.34 体現形の識別子 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 体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}

```

第2図 JPMARC レコード例の変換結果 (部分)