

著者名典拠録の情報構造化記述

研谷 紀夫

東京大学大学院情報学環

〒113-0033 東京都文京区本郷7-3-1

TEL:03-5841-3500

概要

WEB上にある多様な情報の構造化についての議論と実践が行われるようになって久しいが、それらの情報の構造化にあたっては、用語などの統制が重要な鍵を握ることは広く認識されている。これらの中で重要な要素の一つが人名情報の統制である。WEB上には様々な人名情報が存在するが、これらの人名情報を統制し、さらにそれらの関係をより明確化することによって、WEB上の情報を様々なコンテキストで閲覧することが可能となる。これらの人名典拠情報を構築するにあたっては新たに構築するだけでなく、既存の人名情報の活用が最も有効であろう。本研究では、一つのアプローチとして、日本において多年に渡って人名に関する典拠情報を作成してきた国立国会図書館の著者名典拠録をMADS及びMODSなどの情報構造化記述言語での表現とそれらのデータの活用の可能性について論じる。

キーワード: 著者名典拠録、MADS、MODS、メタデータ、JAPAN/MARC(A)、MARC21

Data Structural Description for Name Authority File

Norio Togiya

University of Tokyo

Graduate School of Interfaculty Initiative in Information Studies

7-3-1 Hongou, bunkyo-ku, Tokyo Japan

Abstract

Much research and discussion has been done on structuring the varied and diverse information on the Internet. It is widely accepted that standardization of things such as terminology is an important key to being able to effectively structure such data. An important element in doing so is the standardization of name authority file. Various types of personal name information exist on the web, and by standardizing this personal name authority file and further clarifying the relationships between various pieces of information, it would be possible to browse information on the web in a variety of different contexts. In building this type of personal name authority file, it is most effective to not build such information from scratch, but to instead utilize existing personal name authority file. As one approach for achieving this, this study looks at the name authority file of the Japanese National Diet Library, built from reference information related to personal names in Japan over many years. Then, we use description languages and schema such as MODS and MADS to express this and consider possibilities for its usage.

Keywords: Name Authority File, MADS, MODS, JAPAN/MARC(A), MARC21

1. 研究の背景

国立国会図書館では長年にわたり、著者名典拠情報を蓄積し、近年では館内およびホームページ上の書誌検索だけではなく、近代デジタルライブラリーなどのWEB上のサービスにおいても、該当資料の著作権管理やユーザへの著者名情報の提供などに活用されている。今後は、これらの情報は同館のWEB上のサービスや図書館だけではなく様々な機関やサイトサービスにおいても

活用されると考えられる。

このような活用の際には、現在のJAPAN/MARC形式での格納から、WEB上の情報サービスにおいて活用しやすい、構造化記述言語での記述を行うことが望まれる。そのことによってMARC形式の活用事例のない図書館以外の施設や、海外などの機関においてもこれらの情報を活用する契機となると考えられる。特に日本の著者名などの情報は海外では一部の作家のみに限られる傾向があり、これらの典拠データを海外においても参照することも必要である。そのため、本研究では将来的な活用を視座におきつつ、現在のJAPAN/MARC形式で格納された著者名典拠情報を、WEBサービスでも活用しやすいように、情報構造化をどのように行うべきかに焦点をあてて考察を行う。

現在WEBを中心としたネットワーク上でのデータ交換においては、XMLなどを用いたデータの構造化とその記述がより浸透しつつあり、MARCをはじめとする、図書館におけるデータもMARC形式からXML形式への変換とその対応が進展している。その代表的な例が米国議会図書館によってMARC21をXMLに対応させたMARC XML[1]である。またそれ以外にもMARC XMLよりスキーマの構造を集約しながら設計したMODS(Metadata Object Description Schema)[2]や、典拠情報に関しては、MADS(Metadata Authority Description Schema)[3]などのスキーマを設計している。特に後者のMADSに関しては現在も発展段階にあるため、仕様の更新を行っており、実践例などを積み重ねながら、今後も改訂を行うと考えられる。

一方で、図書館だけではなく文書館などの分野においても、目録・書誌情報及び典拠情報のXMLを用いた構造化記述が浸透しており、典拠情報の基本的なモデルを示したISSR(CPF)の基本構造に具体的なXMLスキーマ定義したEAC-CPFなどをあげることができる[4]。本研究ではこのような動向を受けて、国立国会図書館の協力を得て、同館が作成した「国立国会図書館著者名典拠録」の、XMLなどを用いた構造化記述の試みを行った。これらに関しては段階的にJAPAN/MARC(A)形式のデータを全てXMLで表現した後に、MADS及びMODS形式を用いての記述を行う。

2. JAPAN/MARC (A) 形式の国会図書館人名典拠情報の概要

国立国会図書館が発行している『JAPAN/MARC マニュアル 典拠編 第1版』[5]に示されているように、同館では、1948年より整理部門において典拠カードの作成管理を行ってきた実績がある。さらに1991年には電算機での扱いを考慮した国際フォーマットであるUNIMARC典拠フォーマットの開発を受けて、国会図書館においても典拠情報の電子データ化がより進展し、1997年には『JAPAN/MARC著者名典拠』として、磁気テープ版の典拠情報の刊行が開始された。さらに、2001年にはCD-ROM版である『NDL CD-ROM Line国立国会図書館著者名典拠録 2000年版』が刊行されている。これらのJAPAN/MARC(A)形式の個人名については「001:レコード識別番号」「005:レコード更新情報」「100:一般処理データ」「150:名称に関するコード化データフィールド」「152:目録規則」「200:標目-個人名」「230:標目-統一タイトル」「250:標目-普通件名」「300:名称に関する注記」「320:一般説明参照注記」「330:一般的スコープノート」「400:「を見よ」参照指示-個人名」「686:分類番号-NDLC/NDL」「801:レコード作成機関」「810:出典データ」「830:目録作成者一般注記」「835:標目削除に関する注記」「900:件名参照-個人名」「911:標目区分」「915:官庁コード」「936:標目訂正に関する注記」「980:文字に関する注記」などのブロックがあるが、全てにデータが格納されている訳ではない。著者名典拠録に関しては、主にデータが格納されるのは、「001」「005」「100」「152」「200」「300」「400」「801」「810」「830」「900」「911」「936」などである。

3. JAPAN/MARC (A) 形式のXML形式への変換

MARC形式のXML化に関しては、既に前述したようにMARCXMLという形式が発表されているが、JAPAN/MARC形式のXML化については既に朝倉によってJ-BISCとTRC-MARCへのXML化を実践している[6]。また神崎らによっても実際のデータ例が公開されている[7]。本研

究においてもこれらMARCXMLと朝倉による実践や神崎によるサンプルを基本にJAPAN/MARC(A)形式のXML化を行った。

これらの概要は図1に示される内容である。JAPAN/MARC(A)には、各ブロックナンバーと2つのインジケータが記述されている。その後に各ブロックの下サブフィールドが設定されており、それぞれの値が格納されている。XMLの記述に関しては、前述した実践にならい、各ブロックに<datefield>タグを用いてattributeにブロック番号、インジケータ番号を入力し「<datefield tag="100" ind1="0028" ind2="00026">」ように記述し、ブロック番号と2つのインジケータの値を配置した。また、サブフィールドについてもattributeにサブフィールドの値を入力し、右記のように「<subfield code="a">20001205</subfield>」と記述した。これらの形式は基本的に全てのブロックに継承され、ブロック番号、インジケータとサブフィールド名とその値をXMLの中に格納した。JAPAN/MARC(A)のXML化はその他の記述の仕方も考えられるが、現状のJAPAN/MARC(A)のブロックやサブフィールドを構造化していく上では、これらの形式が最も一般的であると考えられる。

4. MADSとMODSを用いた著者名典拠録の構造化記述

4-1. 構造化記述の概要

前セクションにおいては現在のJAPAN/MARC(A)形式のXML形式の構造化を行ったが、この段階で、より広範囲なデータベースでより共有化しやすい構造となったと言える。しかし、ブロックやサブフィールドの名称は規定の数値や記号などによって表現されており、これらをインターネットなどの情報空間で共有していくためには、より標準的な形式での記述が必要である。前述したように、既に典拠情報は米国議会図書館が策定したMADSの他、文書館においてはEAC-CPFなどが既にこれらの仕様を公開している。本研究では、これらに対応させていく試みとして、JAPAN/MARC(A)形式の典拠データを、MADSとMODSのスキーマを活用して、情報の構造化記述を行った。MODSについてはMARC21とのマッピングなどが公開されている。そのためこれらを参考としながら、JAPAN/MARC(A)とMADS及びMODSとのマッピングを行った。MADSは典拠情報を記述する形式としてスキーマが用意されているが、MODSと同様に、MARCの中のサブフィールドで示されるより詳細なスキーマまでは用意されていない。またMODSの典拠情報を提供する中心的な役割を果たす<name>エレメントも基本的にはMADSにおいて定義されている<name>エレメントの使用法を継承する。またデータ作成やそれらの使用に関する項目もMODSの<recordInfo>を継承して用いることが推奨されており、MODSスキーマを併用しながら、必要な情報を記述することが想定されている。これらの併用関係を想定しながら、JAPAN/MARC(A)形式のデータを、MADSとMODSのスキーマを活用して記述した例の一部を図2に示した。前項で示したように、JAPAN/MARC(A)のブロックの構造を継承するのではなく、それぞれのデータを再構成し、構造化した。

4-2. エレメントデザイン

はじめにMADSによって設定されたスキーマから解説を行う。MADSは主に、基準となる名称などの部分は<authority>エレメントの中に情報を格納し、それ以外の用法や語は<variant>においてその属性を、attributeを使用して定義しながら表現を行う。そのためJAPAN/MARC(A)の「200：標目-個人名」で示されたような情報は<authority>スキーマの中で情報記述を行う。<authority>の中で示される名称はMADSの中で示された<name>エレメントを用いるが、これらの用法はMADSで示された同スキーマの用法と同じであり、これらに基づいた記述を行う。<name>にはattributeが複数用意されているが、JAPAN/MARC(A)のデータの表現においては、主にtype、ID、xml:lang、dateを用いた。JAPAN/MARC(A)においては、名称のデータが複数存在する場合は各名前に「a01」といったIDが付与されている。これらのIDはattributeのIDに記述することとした。またattributeのxml:langについては、各名称を記す言語の種類を記述した。英語圏などの人名典拠情報と日本語の典拠情報で大きく異なる点は、日

本の典拠情報が「漢字」「カタカナ」「英字」の複数の文字の種類によって記されている点である。そのため、最初に対象が日本語での表現であるか、英語での表現であるかをattributeで記すこととした。さらにJAPAN/MARC(A)においては「dc」「ba」にあるように言語の表現がカナであるか漢字であるかを示す値が設定されているが、これらは現段階ではattributeなどを用いてのデータの格納は行わないこととした。

次にattributeのdateにおいては、該当する<name>が使用された時期などを、attributeを用いて記述することができる。そのため、JAPAN/MARC(A)のサブフィールド「f」にあたる生没年の記録は、「<namePart type="date">1882-1934</namePart>」という形式を用いて記述した。

また<name>における姓と名の区別であるが、MADS及びMODSにおいては姓と名が別れる場合は、「<namePart type="family">Yamabe</namePart>」「<namePart type="given">Rozin</namePart>」と分けて記述するか、<namePart>の下に姓と名をカンマなどによって区切ることも考えられる。また、別名などについては<variant>を用いて表現を行うが<name>タグ内における表現は<name>タグと同じである。

次に、該当人物の経歴などに関する情報であるが、これらはMADSにおいては、<note>タグにおける記述が想定されている。<note>においても複数のattributeが用意されているがhistoryが経歴などの情報に該当する。そのため、JAPAN/MARC(A)においては、「830：目録作成者一般注記」ブロックに該当する情報は<note type="history">に格納することとした。本ブロックにおいては、対象人物の職業、職種に関する語が端的に記述されている。そのため、詳細な履歴とは異なるが、それらに該当する情報として、<note type="history">に格納することが最も適切であると考えられる。

また、ブロック810には、該当する著者が記した書籍と履歴情報を記した根拠などが記されている。根拠を記す項目としてはMADSにおける<note type="resource">が最も適切だと考えられる。また根拠を示す書籍などが記されている場合があるが、これは<note type="resource">とともに、関連する書籍のタイトルを表記する、<mods:titleInfo> <mods:title>エレメントにもデータを格納した。

一方でこれらのデータの記録管理に関する情報の格納も必要である。JAPAN/MARC(A)においては、「100 一般処理データ」に該当するものであるが、このような管理情報はMADSにおいてはエレメントは用意されておらず、MODSの<MODS:recordInfo>エレメントを使用することが想定されている。そのため、本稿においてもMODSのエレメントを用いた記述を行う。MODSの<MODS:recordInfo>エレメントには複数のサブエレメントが用意されている。ここで用いたのは、データを作成した機関を示す<MODS:recordContentSource>、データを作成した年月日を表現する<MODS:recordCreationDate>データの更新日を表す<MODS:recordChangeDate>、典拠のIDを示す<MODS:recordIdentifier>、記述にあたって採用した目録記述法を示す<MODS:descriptionStandard>、使用言語を表す「<MODS:languageOfCataloging>」である。それぞれはMARC21などに記述にあたって統制された用語を用いることとする。値の具体例としては「JTNDL」（国立国会図書館）、「20010809」「20010809133100.0」「00000171」「NCR」「jpn」などである。

また、これらに加えて「JAPAN/MARC(A)」にはフィールドとして用意されているが、個人名に関してはデータが格納されていない項目として「686 分類番号-NDLC/NDL」や「911標目区分」で示される件名標目などをあげることができる。これらは現在ではデータが格納されていないものの、将来的に、著者に関するデータを格納することが可能になった場合は、それらに対応する項目として、MODSの分類ターム記入エレメントである<MODS:classification authority="NDLC">も加えた。また同様に件名標目を入力するエレメントとして<MODS:subject authority="NDLH">エレメントも設けた。これらの概要は図2の通りである。

以上のような過程によって、JAPAN/MARC(A)に格納されている著者名典拠情報の多くを格納することが可能となったが、「翻字コード」「追加文字セット」「目録用文字」「外字フィールド」などのサブフィールド部分のMADS及びMODSへの適応は日本語独自の情報内容であり、

```

<?xml version="1.0" encoding="UTF-8" ?>
<!-- ?xml:stylesheet type="text/xsl" href="meta-list.xsl"? -->
<record>
  <leader>00762nx 2200253 45</leader>
  <datefield tag="001" ind1="0009" ind2="00000">00000171</datefield>
  <datefield tag="005" ind1="0017"
ind2="00009">20001205141300.0</datefield>
  <datefield tag="100" ind1="0028" ind2="00026">
  <subfield code="a">20001205</subfield>
  <subfield code="a">a</subfield>
  <subfield code="b">jpn</subfield>
  <subfield code="c">0012</subfield>
  <subfield code="d">da</subfield>
  </datefield>
  <datefield tag="152" ind1="0009" ind2="00054">
  <subfield code="a">NCRT</subfield>
  </datefield>
  <datefield tag="200" ind1="0040" ind2="00063">
  <subfield code="name">1</subfield>
  <subfield code="6">a01</subfield>
  <subfield code="a">谷部</subfield>
  <subfield code="b">路人</subfield>
  <subfield code="f">1 8 8 2 - 1 9 3 4</subfield>
  </datefield>
  <datefield tag="200" ind1="0030" ind2="00103">
  <subfield code="name">1</subfield>
  <subfield code="6">a01</subfield>
  <subfield code="7">dc</subfield>
  <subfield code="a">タニベ</subfield>
  <subfield code="b">ロジン</subfield>
  </datefield>
  <datefield tag="200" ind1="0026" ind2="00133">
  <subfield code="name">1</subfield>
  <subfield code="6">a01</subfield>
  <subfield code="7">ba</subfield>
  <subfield code="a">Tanibe</subfield>
  <subfield code="b">Rozin</subfield>
  </datefield>
  <datefield tag="300" ind1="0009" ind2="00159">
  <subfield code="name">0</subfield>
  <subfield code="a">本名：谷部達人</subfield>
  </datefield>
  =====400 ブロックは省略=====
  <datefield tag="801" ind1="0047" ind2="00416">
  <subfield code="name">0</subfield>
  <subfield code="a">JP</subfield>
  <subfield code="b">National Diet Library, JAPAN</subfield>
  <subfield code="c">20080707</subfield>
  </datefield>
  <datefield tag="810" ind1="0013" ind2="00463">
  <subfield code="a">歌集蘆葉</subfield>
  </datefield>
  <datefield tag="810" ind1="0019" ind2="00476">
  <subfield code="a">根拠：家族回答</subfield>
  </datefield>
  <datefield tag="830" ind1="0057" ind2="00418">
  <subfield code="a">東北大学・名誉教授</subfield>
  </datefield>
  <datefield tag="911" ind1="0012" ind2="00495">
  <subfield code="a">p</subfield>
  <subfield code="b">a</subfield>
  <subfield code="c">s</subfield>
  </datefield>
  <datefield tag="936" ind1="0087" ind2="00507">
  <subfield code="a">標目訂正 ヤベ、ミチト（谷部達人）→タニベ、ロジン（谷部路人）（1 9 9 4 0 1 1 9）</subfield>
  </datefield>
</record>

```

図 1：JAPAN/MARC(A)形式の著者名典拠録の XML への変換（データ内容は個人に関する情報を含むことや諸権利に関する関係から標準的なデータを基に架空のサンプル用データを格納している）

```

<?xml version="1.0" encoding="UTF-8" ?>
<mads:xmlns="http://www.loc.gov/mads/"
xmlns:mods="http://www.loc.gov/mods/v3"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mads/ mads.xsd">
<authority>
  <name type="personal" ID="a01" xml:lang="jp">
  <namePart type="family">谷部</namePart>
  <namePart type="given">路人</namePart>
  <namePart type="date">1882-1934</namePart>
  </name>
  <name type="personal" ID="a01" xml:lang="jp">
  <namePart type="family">タニベ</namePart>
  <namePart type="given">ロジン</namePart>
  <namePart type="date">1882-1934</namePart>
  </name>
  <name type="personal" ID="a01" xml:lang="en">
  <namePart type="family">Tanibe</namePart>
  <namePart type="given">Rozin</namePart>
  <namePart type="date">1882-1934</namePart>
  </name>
</authority>
<variant type="other">
  <name type="personal" ID="a01" xml:lang="jp">
  <namePart>本名：谷部達人</namePart>
  </name>
  <name type="personal" ID="a01" xml:lang="jp">
  <namePart type="family">谷部</namePart>
  <namePart type="given">路人</namePart>
  </name>
  <name type="personal" ID="a01" xml:lang="jp">
  <namePart type="family">ヤベ</namePart>
  <namePart type="given">ロジン</namePart>
  </name>
  =====一部省略=====
  <note type="history" xml:lang="jp">東北大学・名誉教授</note>
  <note type="resource" xml:lang="jp">根拠：家族回答</note>
  <note type="resource" xml:lang="jp">歌集蘆葉</note>
  <mods:titleInfo xml:lang="jp">
  <mods:title xml:lang="jp">歌集蘆葉</mods:titleInfo>
  </mods:title>
  <mods:subject authority="ndlsh" xml:lang="jp">件名番号</mods:subject>
  <mods:classification authority="NDLC">
  <topic xml:lang="jp">分類(仮入力)</topic>
  </mods:classification>
  <mods:subject authority="NDSLH">
  <topic xml:lang="jp">標目(仮入力)</topic>
  </mods:subject>
  <mods:recordInfo>
  <mods:recordContentSource>JTNDL</mods:recordContentSource>
  <mods:recordCreationDate
encoding="marc">20010809</mods:recordCreationDate>
  <mods:recordChangeDate
encoding="iso8601">20010809133100.0</mods:recordChangeDate>
  <mods:recordIdentifier>00000171</mods:recordIdentifier>
  <mods:descriptionStandard
authority="marcdescription">NCR</mods:descriptionStandard>
  <mods:languageOfCataloging>
  <mods:languageTerm
authority="iso639-2b" type="code">jpn</mods:languageTerm>
  <mods:languageTerm authority="iso639-2b"
type="code">en</mods:languageTerm>
  </mods:languageOfCataloging>
  </mods:recordInfo>
</mads>

```

図 2：JAPAN/MARC(A)形式の著者名典拠録の MADS 及び MODS 形式へ変換（データ内容は個人に関する情報を含むことや諸権利に関する関係から標準的なデータを基に架空のサンプル用データを格納している）

どのように格納するかは再検討が必要である。また本考察においてはMARCのインジケータ情報は格納しなかった。これらの扱いを含めて、より標準的な構造化言語の中で既存のMARCに含まれた情報のどの部分を継承し、格納していくかについては検討が必要であろう。

5. 課題と展望

以上のように、本稿ではJAPAN/MARC(A)形式に格納された国立国会図書館著者人名典拠録の人名情報に関する主なデータのMODS及びMADSの表現を行った。英語圏においても共通に使用されるMARC項目については、MADS及びMODSに変換することが可能であるが、日本語独自の様式や、記述者によるより詳細な管理情報などの格納方法も検討が必要である。しかし、用途によっては、JAPAN/MARC(A)形式の全ての情報をXML/RDF形式のMADS及びMADSで表現する必要はなく、WEBなどの情報空間においてどのような用途を用いて、それらを用いるかによってデータの取捨選択を行い、限られた情報の格納で十分であると考えられる。

また、これらのデータ化された情報をRDF/XML形式で出力することも十分に可能である。本稿においては紙幅の関係でRDF/XML記述の具体例を示すことができなかったがMODSのRDF/XML形式の表現は既に神崎らによって実践されている[8]。MADS及びMODSの場合も基本的にresourceを、典拠対象人物として、その人名のIDを示し、RDF/XML上でMADS及びMODSスキーマを用いて対象の著者名典拠情報を記述することが可能であると考えられる。

このような結果を踏まえて、今後の典拠情報の活用を考え場合、WEB情報資源上にあるデータに存在する様々な人名に関する統制や、文化資源情報の検索に活用することが望ましい。そのためこれらに活用するためには、生没年だけではなくその作者と関わりのある標目や分類を格納することがよりその他の情報資源と人名情報を結びつける上で重要であると考えられる。著者は書誌情報を結びつけることによって、これらの人名情報とNDLC及びNDLHを結びつけデータに格納する試験計画を進めており、これらのデータが格納された場合は、より人名情報がその他の知識情報と結びつくことにより、典拠情報の構造化とその活用の有用性がより発揮されえると考えられる。

参考文献

- [1] MARCXML: <http://www.loc.gov/standards/marcxml/> (Accessed:2010-02-10)
- [2] MODS: <http://www.loc.gov/standards/mods/> (Accessed:2010-02-10)
- [3] MADS: <http://www.loc.gov/standards/mads/> (Accessed:2010-02-10)
- [4] EAC-CPF: <http://eac.staatsbibliothek-berlin.de/> (Accessed:2010-02-10)
- [5] 国立国会図書館、JAPAN/MARC マニュアル 典拠編 第1版、2003
http://www.ndl.go.jp/jp/library/data/jmarc_a_manual.html (Accessed:2010-02-10)
- [6] 朝倉秀三、日本語MARCのXMLデータベースへの変換に関する一実験-JAPAN/MARCとTRC MARCの変換、日本図書館情報学会誌、50(4)、pp.159-170、2004
- [7] <http://www.kanzaki.com/works/2007/pub/1129-mods.rdf> (Accessed:2010-02-10)
- [8] 神崎正英、デジタルアーカイブフォーラム(DAF)研究会講演資料、2007、
<http://www.kanzaki.com/works/2007/pub/1129keio.html> (Accessed:2010-02-10)

謝辞

本研究においては国立国会図書館のご協力と内藤求氏のご協力を頂きました。心より御礼を申しあげます。また、本研究は、「科学研究費補助金 若手B ネットワーク文化情報資源で活用する人名典拠情報に関する研究」によって行われました。