BIBFRAMEとその問題点 RDAメタデータの観点から

BIBFRAME and its issues

From the viewpoint of RDA metadata

谷口 祥一¹

TANIGUCHI Shoichi¹

.....

1 慶應義塾大学文学部 (〒108-8345 東京都港区三田2-15-45) Tel: 03-5427-1534 E-mail: taniguchi@z2.keio.jp

1 Faculty of Letters, Keio University (2-15-45 Mita Minato-ku, Tokyo 108-8345)

原稿受理 (2015-01-26)

情報管理 2015, vol. 58, no. 1, p. 20-27. doi: <http://dx.doi.org/10.1241/johokanri.58.20>

著者抄録

新たな図書館目録のメタデータ作成基準・規則である RDA を採用するにあたって、メタデータの構成と構文を規定するメタデータスキーマが必要となる。現在、従来の MARC21 を修正したものが欧米では暫定的に使用されているが、それを置換するものとして BIBFRAME が検討されている。本稿は、現在提案されている BIBFRAME の概要とそれが内包する論点や問題点について論じた。BIBFRAME 提案の背景と経緯をまず簡単に取り上げた後、BIBFRAME の概要を要求定義、概念モデル、メタデータエレメントと語彙の定義、エンコーディング方式に分け解説した。次に、BIBFRAME の問題点と課題を、1) RDA メタデータ作成スキーマの観点、2) 複数コミュニティの共通スキーマの観点、3) 実質的に「リンクする」メタデータの観点のそれぞれから整理し論じた。

キーワード

BIBFRAME, メタデータスキーマ, RDA, FRBR, Linked Data

1. はじめに

筆者は、「RDA (Resource Description and Access) でできることできないこと: RDAの理解に向けて」¹⁾を本誌に執筆した。RDA²⁾とは欧米ですでに採用が始まっている新たな図書館目録の作成基準・規則であり、そのRDAに基づくメタデータを実際に作成・蓄積・提供する際に必要となるメタデータの構成と構文を規定するメタデータスキーマとして、現在検討

されているのがBIBFRAMEである。

後述するように現行MARC21の後継の検討(すなわちRDAメタデータ用のスキーマの検討)がその出発点であったが、BIBFRAMEと名付けられた案の検討段階では、図書館界を超え、出版流通業界作成のメタデータや利用者作成メタデータにも適用できるよう検討の方向性が変更された。

先の記事では、RDAの採用にかかわる事項の1つとしてBIBFRAMEを簡単に取り上げた。しかしながら、

メタデータ作成用スキーマはRDAの有効性をも左右する重要事項であり、現在提案されているBIBFRAMEの概要とそれが内包する論点や問題点について確認しておくことは有益と考えるに至った。BIBFRAMEはいまだ検討段階にあり、その最終形は明らかになっていないが、上記の理由により、あえて現段階で本稿の執筆を試みる。

2. BIBFRAME提案の背景と経緯

図書館目録の変革の起点は、概念モデルを示したFRBR（「書誌レコードの機能要件」、1998年）にある。このモデルに全面的に依拠して新たに策定された図書館目録のメタデータ作成基準・規則が、前述のRDAである（2010年刊行）。RDAは作成すべきメタデータの意味的側面のみを規定し、構文的側面は他に委ねるという立場を取っている。これまでの英米目録規則第2版（AACR2）の立場を踏襲していると理解されるが、その結果、従来MARC21フォーマット等が規定してきた構文的側面の部分が残されることとなった。英語圏における現時点の目録作成の構図は、下記のような状況にある。

- ・ 概念モデル：FRBR, FRAD, 国際目録原則
- ・ レコード／データ項目構成：国際目録原則, MARC21改訂版
- ・ 値の記述規則：RDAと各種細則
- ・ エンコーディング方式：MARC21（ISO 2709）, MARCXMLなど

この構図のうち、レコード／データ項目構成とエンコーディング方式の部分が従来のままであり、米国議会図書館（Library of Congress: LC）などはこれまでのMARC21をRDA用に最低限改訂し、以前と同様なレコードを引き続き作成している。これによって、RDAメタデータを曲がりなりにも実装できたことになるが、暫定的な方式にすぎず、RDAの利点を生かしきれていない。なお、MARC21改訂版に並べて書いた国際目録原則では、旧来の書誌レコードと典拠レ

コードという持ち分けを継承することを宣言しているにとどまり、具体的なデータ項目を規定しているわけではない。

こうした状況下にあって、FRBRを起点とする変革を完成すべく、MARC21を置換する方向に向けた検討が着手された。具体的には米国議会図書館による「書誌フレームワークの変革（Transforming our Bibliographic Framework）」³⁾ イニシアチブが開始され（2011年5月）、既存のMARC21と他のメタデータスキーマ群を比較検討し、加えてLinked Dataの技術を適用することによって、図書館目録メタデータの効率的な作成と最大限の再利用を意図した検討が行われている。背後には、報告書「On the Record」⁴⁾（2008年1月）が示した方向性があり、出版流通機関作成のメタデータ（ONIX [ONline Information eXchange]、その他）を最大限に流用することが意図されている。また、図書館が保有するメタデータをLinked Data形式で広く公開することを推奨するW3C図書館Linked Dataインキュベータグループによる報告書⁵⁾ も背景にあったと推測される。

検討の結果、同イニシアチブからBIBFRAMEという案が2012年11月に公表された⁶⁾。その後、いくつかの仕様文書の公表やデータ例、ツールなどの公開が進められ⁷⁾、併せて各種の機会をとらえて討議の場もたれ、また日常的にはメーリングリスト上で継続して議論がなされている。

3. BIBFRAMEの概要

BIBFRAMEは、いわゆるメタデータスキーマとしてのみとらえるのではなく、それ自身、要求定義や概念モデルを備えた、メタデータエレメントと語彙の定義（メタデータ構成要素定義）およびエンコーディング方式（シリアル化方式）をとらえることが適切である。

1) 要求定義

BIBFRAMEが満たすべき要件を指し、具体的な仕様

設計を進めるうえでの前提と位置づけられるのが、要求定義である。これについて仕様文書にまとめて明記されてはいないが、下記の2つが該当する。

- ・ RDAメタデータの作成・蓄積・提供スキーマとして利用でき、かつ他のコミュニティによるメタデータスキーマともなること。図書館外の多様なメタデータとのリンクや共有を意図し、出版流通業界によるメタデータや利用者作成メタデータなど、さらには図書館目録であってもRDA以外の目録基準・規則に基づくメタデータも含めて受け皿となることを出発点に据えている。つまり、RDAに基づくメタデータのみを対象とはしていないのである。
- ・ Web上での情報流通・共有を意図し、Linked Dataの仕様（4つの原則）に従うこと。Webの世界での潮流であるLinked Dataの方向性に沿ったスキーマとすることを要件としている。

2) 概念モデル

提案されているBIBFRAME案は独自の概念モデルを有している。モデルは、「クラス」と「プロパティ」をもって示されている。対象事象を構成する要素が「クラス」であり、その特徴などの属性およびクラス間の関連を表すものを「プロパティ」と呼ぶ。このような分け方は、セマンティックWebやLinked Dataの基本技術として用いられるRDF (Resource Description Framework) に従ったものである。

BIBFRAMEのモデルは、下記の4つの主たるクラスから構成される⁶⁾。

- 「著作 (Work)」は、RDAやその基盤とされているFRBRが定義する「著作」と基本的に変わるところはない。BIBFRAMEによれば、「目録対象となるアイテムの概念的性質 (conceptual essence) を反映するリソース」と定義され、RDA/FRBRにおける「著作」すなわち個別の知的・芸術的創造物という抽象的な実体に相当する。シェイクスピアによるハムレットなどが典型例となりうるが、他方で単一の情報資源としてし

か存在しない場合も含め、原則すべての対象資源において著作が存在する。

- 「インスタンス (Instance)」は、「著作の個別的かつ物理的な具体化 (embodiment) を反映したリソース」であり、RDA/FRBRの「体現形」に対応する。著作を何らかの媒体に特定の形式・様式で固定化したものであり、図書、雑誌、DVD、Webページなどのかたちで存在する。
- 「典拠 (Authority)」は、RDA/FRBRの個人・家族・団体といった「エージェント (agent)」と、主題表現等に用いるリソース (トピック、場所、時間、分類など) に分かれる。
- 「アノテーション (Annotation)」は、BIBFRAMEの他のリソースに対する付加的情報との位置づけにあり、所蔵情報、図書の表紙画像、情報資源に対するレビューやコメントなどが該当する。

図1に上記のクラスとクラス間の関連を示すプロパティを示した。「典拠」は「エージェント」と主題表現用のものに分けて示し、「アノテーション」もサブクラス「所蔵 (HeldMaterial)」とレビューや内容要旨、目次などに分けて示した。また、クラス「プロバイダ (Provider)」を図中に加えたが、これは出版者や製作者に相当する。各プロパティは意味の方向性をもっており、矢印をもって示した。著作からインスタンスに向けたプロパティは「hasInstance (イ

図1 BIBFRAMEモデル (1)

ンスタンスをもつ)」、逆にインスタンスから著作に向けたプロパティは「instanceOf (インスタンスとなる)」となる。図中の矢印は一方のみの方向性を示しており、それゆえ一部のプロパティは矢印の向きとそぐわない場合もある。それぞれのクラスには、複数の属性（プロパティ）が割り当てられるが、ここでは省略する。

また、BIBFRAMEでは、クラス「タイトル (title)」と「識別子 (identifier)」があり、それにかかわる部分を図2に示した。「著作」と「タイトル」はプロパティ「workTitle (著作のタイトル)」が結びつけ、同様に「インスタンス」と「タイトル」はプロパティ「instanceTitle (インスタンスのタイトル)」「abbreviatedTitle (短縮形タイトル)」などが関連となる。図2には、クラス「著作」と「エージェント」の属性に当たるプロパティ「authorizedAccessPoint (典拠形アクセスポイント)」のみあえて示した。

ここで取り上げたものの以外にもBIBFRAMEは多数のクラスとプロパティの定義を示しており、それらを含め、さらに展開したモデルとすることもできる。

図3に、FRBR報告書を対象資源とした事例のモデル図を示した。BIBFRAMEの文書に例示されているものを借用したが、現在の仕様案に合うよう一部変更している。単一の著作が3つのインスタンスと関連づけられ、それぞれのインスタンスがその出版者と関連づけられている。加えて、著作はその作成者と主題

図3 BIBFRAMEモデル (事例の例示)

に相当するトピックとつながられている。

3) メタデータエレメントと語彙の定義

先のモデルに対応するよう、メタデータスキーマを構成するエレメント（記録項目）の定義と語彙の定義を、RDFクラスとプロパティをもって定めている。先の概念モデルは、このクラスとプロパティの定義から抽出し構成したものである。

メタデータの構成要素であるエレメントが「プロパティ」として定義され、対象とするリソースの属性に該当する。プロパティはその主語となるリソース（対象となるリソース）と目的語となるリソースがそれぞれ属するカテゴリを「クラス」を用いて指定しており、主語となるリソースのクラスを「定義域（ドメイン）」、目的語となるリソースのクラスを「値域（レンジ）」と呼ぶ。たとえば、図1のプロパティ「hasInstance」の定義域はクラス「著作」、値域は「インスタンス」とされており、逆に「instanceOf」の定義域は「インスタンス」、値域は「著作」となる。同様に、図2に出現したプロパティ「workTitle」の定義域は「著作」、値域は「タイトル」である。他方、値域がリテラル（文字列など、他のリソースにつながらないもの）をとるプロパティも多数ある。「titleAttribute (タイトルの属性)」「titleQualifier (タイトルの限定句)」「titleType (タイトルのタイプ)」「titleValue (タイトル文字列)」などは、定義域が「タイトル」、値域がリテラルをとる。

図2 BIBFRAMEモデル (2)

BIBFRAMEは固有のクラスとプロパティを定める一方、他所で定義されたものも組み合わせてそのまま使用できとしている。

4) エンコーディング方式

最後に、メタデータをエンコーディング（シリアル化）する方式、すなわちコンピューター処理可能な形式とする記録化方式には、RDFデータをXML（Extensible Markup Language）でエンコーディングするRDF/XMLや、同データのJSON（JavaScript Object Notation）を用いたエンコーディングなどが予定されている。各種の方式でエンコーディングされたメタデータ間に意味上の差異はなく、今後も新たな方式が追加される可能性がある。

4. 現段階のBIBFRAME案の問題点と課題

BIBFRAMEの現段階の公表案について、ここで複数の観点からその問題点等を検討しておきたい。

1) RDAメタデータ作成スキーマとして見たときの問題点

BIBFRAMEは前述のとおり、図書館外の多様なメタデータへの適用を意図しており、RDAに基づくメタデータのみを対象とはしていない。そこから、RDA/FRBRとは異なる独自の概念モデルを採用することになった。換言すれば、RDA/FRBRモデルでは、他のコミュニティのメタデータ（出版流通業界のメタデータや利用者作成メタデータなど）、あるいはRDA以外の目録基準・規則に基づくメタデータには当てはめられないとの事情がある。RDAそしてFRBRは対象とする情報資源を「著作」から「個別資料」まで4つの実体（クラスに相当）をもってモデル化する（図4）。ここでの「表現形」は、著作を文字・数字、画像等によって表現した部分であり（例：ハムレットの原語版、日本語訳など、さらには福田恆存による日本語訳など）、音楽著作の場合には個々の演奏が表現形に該当する。個別資料はその1点ずつのコピー（図書の場合の複本の1点ずつ）を指す。

図4 RDAモデル（FRBRに相当する部分）

前述のとおり、BIBFRAMEとRDA/FRBRの両者の「著作」は対応づけられ、またBIBFRAMEの「インスタンス」とRDA/FRBRの「体現形」が対応づけられる。ここからRDAの「表現形」の扱いが問題となる。この点は案の提案当初から議論となり、著作に組み入れるべきか、インスタンスの一部とすべきか見解が分かれた。現時点では、プロパティ「hasExpression（表現形をもつ）」「expressionOf（表現形となる）」の定義域と値域がともに「著作」とされている点、BIBFRAMEプロファイルの仕様書においてRDAとの関係を示した部分でRDAの著作と表現形をとともにBIBFRAMEの著作に対応づけている点をもって、RDAの表現形はBIBFRAMEの著作の一部と見なすことに落ち着いたとみられる。この結果、RDAで単一著作の異なる表現形としたものは、BIBFRAMEでは表現形ごとに新たな著作となる。たとえば、RDAで1つの著作とその2つの表現形とされたものは、BIBFRAMEでは3つの著作として表されることになる。RDAにおいても表現形は確かに扱いが堅固とはいえず、いくぶん希薄な位置づけではあるが、BIBFRAMEでの扱いとは大いに異なる。

同様に、BIBFRAMEプロファイル仕様書では、RDAの個別資料は体現形とともにBIBFRAMEのインスタンスに対応づけられているが、アノテーションの一種とされる「所蔵」やそのサブクラス「アイテムの所

蔵 (HeldItem)」とも対応する部分があり、その持ち分けには曖昧さがある。

また、属性／プロパティの割り当てにおけるRDA/FRBRとの齟齬がある。具体的には、RDA/FRBRでは著作／表現形の属性であるものが、BIBFRAMEの「インスタンス」のプロパティとされている例がある。プロパティ「contentsNote (内容注記)」「awardNote (受賞注記)」「colorContent (色量)」などが、これに該当する。

以上のことにより、BIBFRAMEはRDAメタデータを素直に表現できるスキーマとは考えにくく、RDAメタデータの「作成用」スキーマとしてみたときには問題がある。いったんBIBFRAMEスキーマで作成されたメタデータはRDAの構造に復元することは困難となる。なお、メタデータ作成用ではなく、外部への提供用スキーマであれば上記の点は問題視する必要はない。

2) 複数コミュニティにまたがる共通スキーマとしてみたときの課題

BIBFRAMEモデルが情報資源を著作とインスタンスという2つのクラスでとらえていることを、次に複数コミュニティにまたがる共通スキーマの観点からみておきたい。これら2つのクラスはおおむね「内容 (content)」と「キャリア (carrier)」の区分を意図したものととらえることができる。ただし、こうした区分や具体的な線引きはコミュニティにより相違がある。異なる考え方や基準に基づき作成された個々のメタデータをBIBFRAMEの「著作」と「インスタンス」にどこまで統一したかたちで分解できるのか、分解したうえで同一著作、同一インスタンスを同定し統合できるのか定かではない。著作の認定の問題を含めて、ここには困難さが伴う。

ここで強調しておきたいのは、個々のコミュニティにおける要求事項には差異があり、それゆえ概念モデルそしてメタデータスキーマがコミュニティごとに（さらには機関ごとに）策定され使用されている点である。文書館、博物館・美術館

において使用されているものは、図書館において使用されるものとはおのずと異なる。換言すれば、BIBFRAMEをメタデータ「作成用」スキーマに採用するコミュニティは図書館以外にはおそらくありえない。こうした中で、なぜ図書館だけがRDAとも異なる、つまりRDAメタデータを素直に表現できないスキーマをあえてRDAメタデータ作成用に使用すべきなのであろうか。この点に大きな疑問を抱く。これは、メタデータ提供用スキーマを他コミュニティとの共有スキーマとすること、つまり必要な時点でBIBFRAMEスキーマにメタデータを変換して提供することとは、本質的に異なる問題である。

3) 実質的に「リンクする」メタデータとするための課題

BIBFRAMEの要求定義にも含まれているとおり、BIBFRAMEはWeb上での情報流通・共有を意図し、Linked Dataの仕様に従っている。Linked Dataとは、コンピューター処理に適したデータをWeb上で公開・共有する仕組みとその行為の総称であり、広く公開されたデータをLinked Open Data (LOD) と呼ぶ。従来のWebがHTML文書間のハイパーリンクによって主に人間向きに情報空間を構築してきたことに対して、Linked Dataは構造化されたデータ同士をリンクさせることでコンピューターが処理可能な「データのWeb」の構築を目指している。

このようなLinked Dataに向けた展開の方向性は妥当であろう。Webの世界での潮流であるLinked Dataの動きに同調し、図書館が保有するメタデータの公開を一段と進めていくことは大いに推奨される。図書館とそのサービスの有用性を広く認知してもらう格好の機会であろう。

他方、他のメタデータと実際につながり機能するLinked Dataとするには、あらゆる事物（リソース）にURIを付与するなどの示された原則に加えて、[1] クラス内の個々の値となる個別の著作、インスタンス、エージェント等に、ローカルなデータベースの範囲を越えて共通のURI（一元化されたURI）を付与

することが必要となる。あるいはそれが困難であれば、[2] 事後的に事物の一致が判定できるような記述が求められる。それぞれに付与されたURIが同一であれば同一リソースとなるが、そうでなければ記録された他の手がかりによって事後的に同一性を判定せざるを得ない。[1] についてはそれを支える新たな大きな仕組みが必要であり、関連するコミュニティーや団体との識別子をめぐる交渉と調整となる。[2] については、Linked Dataでは「レコード」単位から個別エレメントの記録に相当する「トリプル」単位でのメタデータ流通が強調されることが多い。つまり、レコードというまとまりを必要としない（固定しない）考え方である。これまではレコードを構成する複数項目の組み合わせによって、リソースの同一性判定を行ってきたが、それが担保されない「トリプル」単位では同一性の判定は困難となろう。Linked Open Dataとして公開されたものを、属性値である名称文字列が一致したら同一リソースと見なすなど、利用者が任意の方式で同定すること自体には問題はない。それぞれの考えと責任において同一性を判定しリンクさせることに特段の不都合は生じない。しかし、図書館が同一性を判定しリンクさせる場合には、多少なりとも慎重さが求められる。

ちなみに、RDAメタデータを素直に表現できるスキーマを想定したとき、併せてLinked Dataの原則

に沿ったスキーマとすることは可能である。Linked Dataは、基本的に提供用スキーマの要件であって、メタデータ作成処理やその作成用スキーマの在り方に変更を及ぼすものとは考えにくい。

5. おわりに

本稿ではBIBFRAMEの現段階での案を紹介したうえで、その内在する問題や課題、特にRDAメタデータスキーマの観点からみたときの問題点等を論じた。

筆者の主張は、RDAを素直に表現できるスキーマによってRDAメタデータを作成すべきであり、それを必要な時点でBIBFRAMEスキーマによるメタデータに変換して提供する、つまりBIBFRAMEは他コミュニティーのメタデータとの共有スキーマとしてのみ利用すべきというものである。今後、議論の新たな展開によって、上記のような方向に進むことを少しでも期待したい。

谷口 祥一（たにぐち しょういち）

図書館情報学の領域において、主に情報組織化、情報資源組織化にかかわる教育と研究に取り組んでいる。特にメタデータ、分類、統制語彙／オントロジーとそれらの組み合わせによる情報と情報資源の組織化に関心がある。図書館情報大学卒。2012年3月まで筑波大学に在籍し、その後は現職。

参考文献

- 1) 谷口祥一. RDA (Resource Description and Access) でできることできないこと：RDAの理解に向けて. 情報管理. 2014, vol. 56, no. 11, p. 758-765.
- 2) RDA Toolkit. <http://www.rdatoolkit.org/>, (accessed 2014-11-01).
- 3) Library of Congress. "Transforming our bibliographic framework". 2011-05-13. <http://www.loc.gov/bibframe/news/framework-051311.html>, (accessed 2014-11-01).
- 4) The Library of Congress Working Group on the Future of Bibliographic Control. "On the record". 2008-01-09. <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>, (accessed 2014-11-01).
- 5) Library Linked Data Incubator Group. "Final report". 2011-10-25. <http://www.w3.org/2005/Incubator/ld/>

XGR-Ild-20111025/, (accessed 2014-11-01).

- 6) Library of Congress. "Bibliographic framework as a Web of data: Linked Data model and supporting services". 2012-11-21. <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>, (accessed 2014-11-01). [日本語訳]: "データのウェブとしての書誌フレームワーク：リンクトデータ・モデルと支援サービス". http://www.nii.ac.jp/CAT-ILL/archive/pdf/Bibliographic_Framework_as_a_Linked_Data_Model_Translation.pdf, (accessed 2014-11-01).
- 7) BIBFRAME.org. <http://bibframe.org/>, (accessed 2014-11-01).

Author Abstract

Metadata schema that defines the structure and syntax of metadata is required to implement RDA, which is a new set of guidelines and instructions on formulating metadata in library catalogs. While the revised MARC21 formats have been used provisionally as such metadata schema for RDA in western countries, BIBFRAME has been proposed and examined to replace MARC21. This paper discusses the outline of the draft of BIBFRAME and issues involved in it. First, the background and circumstances of BIBFRAME are briefly described, and then its functional requirements, conceptual model, metadata elements and vocabulary definition, and encoding schema are explained. Next, issues and problems involved in BIBFRAME are discussed from the viewpoint of 1) schema for RDA metadata creation, 2) schema for sharing various metadata in different communities, and 3) metadata to be "linked" substantially.

Key words

BIBFRAME, metadata schema, RDA, FRBR, Linked Data