

RDA と NCR2018 にとって適切なメタデータスキーマとは何か

谷口 祥一 (慶應義塾大学文学部)
taniguchi@z2.keio.jp

[抄録] RDA や NCR2018 を適用してメタデータを作成する際に用いるべき適切なメタデータスキーマとはどのようなものか検討した。現時点で候補となる MARC21、RDA Registry の登録語彙 (RDA 語彙) を用いたスキーマ、BIBFRAME、および MARC21 から他の 2 つへの事後的な変換という 5 つの方式それぞれに対して、別途設定した複数の観点から検討し、各方式の特徴や問題点を明らかにした。加えて、RDA 語彙において欠如している点を補うべく、構造化して値を表現できるよう拡張を提案した。

1. はじめに

RDA や近々完成予定の NCR2018 はあくまでもメタデータの記述規則というスタンスを取っており、それらを適用してメタデータを作成する際のスキーマについては規定していない。現在 RDA を採用している機関の殆どは、従来からの MARC21 フォーマットを RDA 用に最小限変更したスキーマを用いてメタデータ作成を行っている状況にある。NCR2018 も当初は、同様の方式で使用されるものと予想される。他方で、これはあくまでも移行措置として捉えられており、LC 等は BIBFRAME という、RDA に限定されない多様なメタデータの交換・共有スキーマを提案し、それを RDA メタデータに適用すべく検討を進めている。

本発表は、RDA や NCR2018 を適用してメタデータを作成する際に用いるべき適切なメタデータスキーマとはどのようなものかを検討することを目的とする。具体的には、①現時点で候補となる MARC21、RDA Registry¹⁾ の登録語彙 (RDA 語彙) を用いたスキーマ、BIBFRAME²⁾、および MARC21 から他の 2 つへの事後的な変換という 5 つの方式に対して、設定した複数の観点から検討し評価を試みる。加えて、②候補の一つである RDA 語彙に対して構造化して値を表現できるよう拡張を提案する。なお、RDA エlement と NCR2018 エlement は 1 対 1 対応であると想定し、これ以降は NCR2018 を RDA と区別して取り上げることせず、RDA のみで検討を進める。

RDA 語彙は、RDA Steering Committee が維持管理しており、RDA エlement、関連指示子、値に用いる語彙などが包括的に登録されている。Element と関連指示子は RDF プロパティ、値語彙は RDF クラスとして URI を伴って登録されている。現時点では、バージョン 2.7.3 が最新であるが、並行して IFLA Library Reference Model への対応などを意図

した、3R プロジェクトが進行中であり、それに対応する暫定バージョン 3.0 が GitHub で公開されており、流動的である。しかしながら、これら語彙の使用目的は明言されておらず、メタデータ作成スキーマとしての使用を意図しているか不明である。

他方、BIBFRAME は LC が維持管理しており、現在、バージョン 2.0 (2016.04 公開) であるが、その後も微修正が行われている。RDF のクラスとプロパティをもって定義しており、独自の概念モデルに基づいている。RDA メタデータ以外にも、多様なメタデータの受け皿すなわち交換・共有スキーマとして策定されている。LD4L による bibliotek-o という拡張の検討と提案もある。

2. RDA メタデータ作成方式の類型化および検討・評価の観点

RDA を採用しつつ、メタデータ作成スキーマに MARC21 を用いている現行方式 (方式 1) に加えて、そこから公開・提供スキーマとして事後的に RDA 語彙によるスキーマに変換する方式 2、同様に BIBFRAME に変換する方式 3 を設定した。加えて、作成から公開・提供までを RDA 語彙によるスキーマを用いる方式 4、同様に BIBFRAME を用いる方式 5 を設定した (表 1)。なお、作成用スキーマはそのままメタデータの永続的蓄積用のスキーマになると想定し、それに対し公開・提供用スキーマは区別している。

これらスキーマおよび方式を検討・評価する際の観点を、以下のように設定した。

観点 1: RDA が定める Element や値を過不足なく表現できるか。RDA Element を選択的に採用すること (特にコア Element 以外) は認められており、LC-PCC などの適用細則においても行われているが、ここでは RDA のすべての Element 等を用いることを想定する。

表 1 RDA メタデータ作成方式

	概念モデル	メタデータ作成・蓄積スキーマ	公開・提供スキーマ
方式 1 (現行)	パリ原則に基づくモデル	MARC21 ↑ RDA エlementからの マッピング	同左
方式 2	同上 (+FRBR 等?)	同上	→ RDA 語彙によるスキーマに 変換
方式 3	同上 (+BIBFRAME?)	同上	→ BIBFRAME に変換
方式 4	FRBR 等	RDA 語彙によるスキーマ (=RDA エlement)	同左 (→ 必要であれば、 BIBFRAME に変換)
方式 5	BIBFRAME	BIBFRAME ↑ RDA エlementからの マッピング	同左

この観点とは、作成スキーマに対する問題であり、公開・提供スキーマのみでの対処は困難である。
観点 2: エlement間との関係とその値を適切に表現できるか。複数の並列タイトル、タイトル関連情報、並列タイトル関連情報などの対応関係、あるいは複数の出版地、出版者、出版年の対応関係など、複数のElement間との関係や値の対応関係を適切に示せるか、また値とその読みなど、構造化して表現できるかという観点。これは、作成スキーマと公開・提供スキーマの両者にかかわる問題である。

観点 3: 作成されたメタデータが「著作—表現形—体現形—個別資料 (W-E-M-I)」(またはそれに相当するクラス)の構成をもって表現され活用されるか。ここには、上記のクラスへのメタデータ分割と照合・統合が行えることが含意される。この観点は基本的に作成スキーマの問題であるが、公開・提供スキーマでの機械的変換の問題も伴う。

観点 4: Linked Data に沿ったものとなるか。具体的には、RDF と URI を用いてスキーマが構成されているか。これは、主に公開・提供スキーマの問題である。

また、追加的な観点として下記を設定した。
観点 5: 既存メタデータ(レコード)への対処、システム資産の改修の程度。これは作成スキーマの問題である。

観点 6: 他のメタデータモデルや語彙との相互運用性、すなわちマッピングの容易性など。公開・提供スキーマの問題である。

3. 各方式の検討と評価

3. 1 方式 1

現在、採用されている方式であり、図書館目録の既存メタデータは、以前の AACR を適用したものを含めて、すべてこの MARC21 によって表現されている。それゆえ、観点 5 からは

追加的な対処は不用である。

観点 1: RDA Steering Committee 自身による RDA エlementから MARC21 フィールド/サブフィールドへのマッピングが公表されており、原則すべてのElementがマッピングされる。ただし、それは多対一のマッピングとなる場合を含む。例えば、Parallel Title Proper、Other Title Information、Parallel Other Title Information は、すべて MARC21 の「245 b」にマッピングされる。これらゆえ、RDA エlementとして機能しない。

観点 2: MARC21 サブフィールドの値同士の対応関係は示せない。値とその読みなどについては、変則的な扱いでの対応づけである。

観点 3: W-E-M-I という構成ではなく、パリ原則に基づく構成である、「統一タイトル典拠レコード—書誌レコード—所蔵レコード」、加えて名称典拠レコードと件名典拠レコードからなる。事後的に W 単位などにクラスタ化する方策もあるが、大きな限界がある。

観点 4: MARC21 フィールド、サブフィールドなどが、URI を付して Open Metadata Registry³⁾に登録されており、RDF データ化が可能である。フィールド、サブフィールドに対応するプロパティは、その定義域と値域の指定はないが、実質的に定義域は各レコードが表すリソース(例えば、「書誌レコードが表すリソース」)のクラスとなり、W-E-M-I などのクラスではない。

観点 6: MARC21 は図書館目録に特化したものであり、他のモデルや語彙からのマッピングは現実的ではない。

3. 2 方式 2

MARC21 により記録されたメタデータを事後的に RDA 語彙によるスキーマに変換する方式である。

観点 1: RDA Steering Committee による MARC21 フィールド/サブフィールドから RDA エlementへのマッピングがあるが、一対多の対応づけとなる部分がある。これゆえ、最終的には「RDA→MARC21→RDA」というマッピングの連鎖となり、RDA エlementを適切に表現していることにはならない。

観点 2: RDA 語彙のプロパティは、定義域、値域ともに未指定の **unconstrained property** と、定義域の指定あり、値域の指定なしという **canonical property** の両方が登録されている。本研究では後者のみ扱うことにするが、それらの定義域は **W-E-M-I** や **Agent** などであるため、それらを組み合わせてElement間の関係とその値を適切に表現したり、構造化した値を表現したりすることはできない。さらに、元となる MARC21 の構造表現力の制約もある。

観点 3: 書誌レコードからの **WEMI** への分割はできるが、分割後の事後的な（機械的な）照合・統合にとどまるため、精度に限界がある。つまり、メタデータ作成時に、既存の著作・表現形との人手による同一性の確認は行われていない。他のリソースに包含される（から抽出された）著作・表現形との同一性の確認は行われず、他の著作・表現形との関連付けも行われない。なお、人手により統一タイトル典拠レコードとのリンクが形成されている場合を除く。

観点 4: RDA 語彙によるスキーマに変換した時点で **Linked Data** に沿ったものとなる。

観点 6: 他のモデルや語彙から RDA 語彙へのマッピングとなるため、現実的ではない。

3.3 方式3

MARC21 により記録されたメタデータを事後的に **BIBFRAME** に変換する方式である。

観点 1: LC による MARC21 フィールド/サブフィールドから **BIBFRAME** プロパティへのマッピングが公表されている。マッピングはリテラル値の記載を組み合わせで一対一となるよう設定されているが、他方ですべてのサブフィールドがマッピングされるわけではない。全体としては「RDA→MARC21→BIBFRAME」というマッピングの連鎖となり、RDA エlementを適切に表現していることにならない。

観点 2: **BIBFRAME** は複数の値の対応関係を適切に示すことができ、また構造化した値を表現することができるが、メタデータが MARC21 で記録されている時点で既に限界があるため、**BIBFRAME** のもつ表現力を活かすことができないと考えられる。

観点 3: RDA の表現形が **BIBFRAME** では著作の一部として包含されており、適切に扱えない。また、方式 2 の場合と同様、書誌レコードからの「著作-インスタンス (Instance)」への分割はできるが、分割後の事後的な照合・統合にとどまり、メタデータ作成時に、著作・表現形の手による同一性の確認は行われない。他のリソースから抽出された著作・表現形との同一性の確認は行われず、また他の著作・表現形との関連付けも行われない。ただし、統一タイトル典拠レコードにより記録された著作と表現形は、人手による同一性の判定が行われており、それとのリンクがある場合を除く。

観点 4: **BIBFRAME** に変換した時点で **Linked Data** に沿ったものとなる。

観点 6: 他のモデルや語彙から **BIBFRAME** へのマッピングや変換は比較的容易である。

3.4 方式4

作成から公開・提供まで RDA 語彙によるスキーマを用いる方式である。

観点 1: 当然ながら RDA エlementとその値を過不足なく表現できる。

観点 2: 方式 2 において述べた通り、現行の RDA 語彙は、Element間の関係とそれらの値を適切に表現したり、構造化した値を表現したりすることはできない。

観点 3: **W-E-M-I** の構成をもって表現され活用される。手元の個別資料から著作、表現形、体現形を抽出する際、すなわちメタデータ作成時に、既登録の著作・表現形・体現形との同定が人手によって行われる。なお、RDA が規定する通り、すべての個別資料から著作、表現形が抽出されるわけではない。

観点 5: 既存メタデータから機械的な変換が必要となるが、変換し抽出された著作や表現形などの照合・同定に限界がある。

観点 6: 他のモデルや語彙に基づくメタデータを、RDA 語彙のスキーマに変換することは現実的ではない。他方、事後的に公開・提供用にさらに **BIBFRAME** に変換するという選択肢はある。これによって、多様なメタデータをすべて **BIBFRAME** に変換し、統合して公開・提供を図ることも可能である。

3.5 方式5

作成から公開・提供まで **BIBFRAME** を用いる方式である。

観点 1: RDA のElementに対応しておらず、RDA エlementと一対一、多対一、一対多、

多対多という対応づけが混在している⁴⁾⁶⁾。また、RDA エlementに対応するプロパティがないケースがある。これらゆえ、プロパティの追加が必要となるが、対応づけ可能な既登録のプロパティとの整合性など、単純ではない。

観点2: BIBFRAME が有する表現力の範囲で、Elementや値の対応関係などを表現できる。こうした表現力は、他のスキーマに比べて高い。

観点3: RDA の表現形を適切に扱うことは困難である。また、BIBFRAME の著作、インスタンス、個別資料というクラスは disjoint ではなく、プロパティの定義域はこれら複数のクラスの union を指定したものが大半である。RDA メタデータに適用する範囲では、仕様で disjoint の扱いにするという選択肢はあるが、それが機能するのかわ不明瞭である。

観点5: 既存メタデータから機械的な変換が必要となるが、変換し抽出された著作(と表現形)の照合・同定に限界がある。

観点6: 他のモデルや語彙に基づくメタデータを変換し、受け入れることは比較的容易である。

4. RDA 語彙の拡張案の提示

各方式の比較検討結果を踏まえて、RDA 語彙によるスキーマにおいて欠如している点(観点2からの評価)を補うべく、構造化して値を表現できるよう拡張を試みる。ただし、RDA 語彙自体が流動的であるため、確定した提案とはしにくい状況にある。下記は、可能な拡張法の1つである。

・RDA Element (Elementサブタイプ、サブElementは除く)に対応するプロパティは、オブジェクト型プロパティとする。

rdam:P30134 "has title" や rdam:P30111 "has publication statement" などが該当する。その値域には、クラス Title や PublicationStatement など、新規に追加したクラスを指定する。

・RDA ElementサブタイプおよびサブElementに対応するプロパティの定義域を WEMI などから、新規に追加する上記のクラスに変更する。

・タイトル関連情報と並列タイトル関連情報などの組み合わせ、出版地と並列出版地、出版者と並列出版者などの組み合わせ、あるいは出版地・出版者・出版日付の組み合わせごとに空白ノードをもって構造化を図る。

図1に出版表示への拡張案の適用例を示す。

引用文献

- 1) RDA Registry. <http://www.rdaregistry.info/>
- 2) Library of Congress. *BIBFRAME Model, Vocabulary, Guidelines, Examples, Notes, Analyses.* <http://www.loc.gov/bibframe/docs/index.html>
- 3) Open Metadata Registry. <http://metadataregistry.org/>
- 4) Shoichi Taniguchi. Is BIBFRAME 2.0 a suitable schema for exchanging and sharing diverse descriptive metadata about bibliographic resources? *Cataloging & Classification Quarterly*. Vol.56, No.1, 2018, p.40-61.
- 5) Shoichi Taniguchi. Examining BIBFRAME 2.0 from the viewpoint of RDA metadata schema. *Cataloging & Classification Quarterly*. Vol.55, No.6, 2017, p.387-412.
- 6) 谷口祥一. BIBFRAME とその問題点: RDA メタデータの観点から. *情報管理*. Vol.58, No.1, 2015, p.20-27

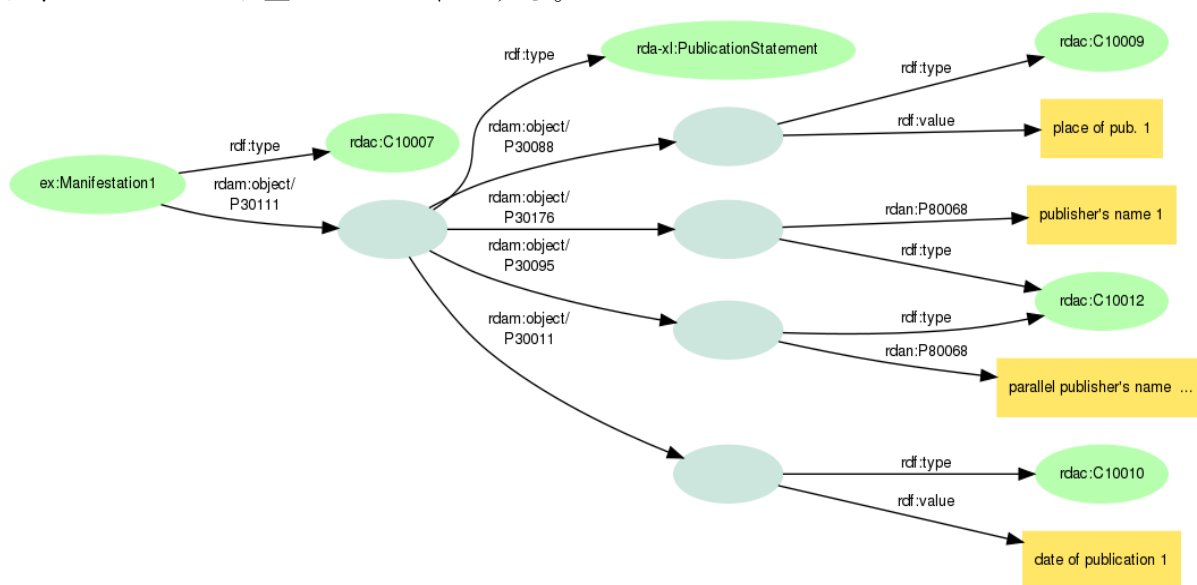


図1 RDA 語彙への拡張適用例 (出版表示への適用)