

RDA から IFLA LRM/DC/MARC21 へのマッピングとその複雑性指標

谷口 祥一 (元慶應義塾大学)

s.taniguchi@keio.jp

RDA エlementから IFLA LRM の属性・関連、DC (ダブリンコア) ターム、MARC21 書誌・典拠フォーマットへのマッピングが公表されている。いずれも規模が大きく複雑であり、全体の把握は困難な状況にある。本研究は、それぞれのマッピングに対して、a) 基本的な集計を行い、次に b) RDA エlementベースの複雑性指標と c) エlement階層木ベースの複雑性指標を提案し計算した。これらによって個々のマッピングの特徴把握が一定程度実行できた。

1. はじめに

RDA 運営委員会は最新の RDA から複数の語彙やスキーマへのマッピングを策定し、RDA Registry¹⁾上で公開している。その主たるものに、RDA エlementから、a) IFLA LRM (図書館参照モデル) の属性・関連、b) DC (ダブリンコア) メタデータ語彙のターム (DCMI metadata terms)、c) MARC21 書誌・典拠フォーマットへのマッピングがある。入念に作成されたものとはいえ、RDA エlement数の膨大さに起因してマッピング全体の規模が大きく複雑である。また、マッピングの策定方針などは明示されておらず、メタデータ変換ツールへの実装を意図したものとも考えにくい。これらゆえ、全体の把握は困難と言わざるをえない。本研究は、前回発表した MARC21 へのマッピングを含めて²⁾、改めて RDA から複数のスキーマへのマッピングを取り上げ、それぞれの全体像を提示し、その複雑性を適切に表す指標群を提案し計算する。

2. 基本的な特徴と集計結果

前回使用した RDA 語彙 v5.1.0 (2024 年 7 月) におけるマッピングを引き続き使用した。このマッピングの特徴は、a) マッピング先要素はスキーマに依存して 1 または多となる、b) RDA エlementには階層関係 (ただし、上位Elementが複数ある多重階層) が定義されている、c) マッピング先の要素にも階層関係・包含関係があることなどである。

1) LRM へのマッピング : LRM の属性と関連は、IFLA によって RDF プロパティとして定義されており、RDA からのマッピングもこれらを用いている。例えば、「rdam:P30156 rdfs:subPropertyOf lrmer:R13 .」として、rdam:P30156 (title proper) は lrmer:R13 (has appellation) にマッピングされている。マッピング元の RDA エlement数 (異なり数) は 3,024、マッピング先の LRM プロパティ数は 103 であり、これらのマッピングペアは

3,024 であった (表 1)。RDA エlementはすべて単一の LRM プロパティにマッピングされている。

2) DC へのマッピング : DC タームへのマッピングも、「rdam:P30156 rdfs:subPropertyOf dct:title .」のように指定しており、rdam:P30156 は dct:title にマッピングしている。出現した RDA エlement数 1,153、DC ターム数 33、またマッピングペア数 1,155 であった。マッピング先を 2 つの DC タームとする RDA エlementが 2 つのみあり、いずれも同じ DC タームの組み合わせであった。

3) MARC へのマッピング : MARC21 書誌フォーマットと典拠フォーマットへのマッピングを統合して扱い、固定長フィールドと可変長フィールドの両者を含めて扱う。例えば、「rdam:P30156 rdakit:hasM21 "M21 Bib 245 ** \$a [unstructured description]" .」は、RDA エlement rdam:P30156 の「非構造記述 (unstructured description)」の値は、MARC21 書誌フォーマットのフィールド 245 (Title Statement)、サブフィールド a (Title) に対応づけることを表している (第 1・2 インディケータの「*」は任意の値を表す)。RDA では 4 つの「記録方法」があり、本研究では「Element + 記録方法」をマッピング元の単位とし、以降はこれを便宜的に「RDA エlement」と呼ぶ (例 : rdam:P30156 [unstructured description])。また、マッピング先 MARC 可変長フィールドは、第 1・2 インディケータとサブフィールドを組み合わせたものを、以降では便宜的に「MARC フィールド」と呼ぶ (例 : B245 ** \$a)。MARC21 書誌の場合は接頭辞「B」、典拠の場合には「A」を付加して区別する。そこでは、「B245 ** \$a, c, p」や「B710 ** \$a, b, c, d, f, g, k, l, n, p, t, u (if subfield \$t is used)」といったマッピング先も登場する。他方、「B245 ** \$7(dpecou)」など、data provenance サブフィールドと名づけられた、値の由来の記録に該当するマッピングは、本研

究の対象外としてすべて除外した。

出現した RDA エlement 数 8,258 (記録方法を組み合わせない、Element のみの場合は 2,481)、MARC フィールド数 1,259、ペア数 34,228 であった。RDA Element ごとに平均 4.1 (SD 2.8) ペアをもち、最大は 30 ペアであった。可能性のある MARC フィールドを幅広く網羅した結果である。

3. RDA Element ベースの指標

先の特徴を踏まえたマッピングの複雑性を表す指標の 1 つは、RDA Element ベースの指標とし、RDA Element 階層と整合したマッピング先の割合を表す指標群を採用した。

RDA Element は階層構造をもち、`rdfs:subPropertyOf` によって定義される。ただし、単一 Element が複数の上位 Element をもつ多重階層である。これは定義域とするクラスの階層 (agent > collective agent > corporate body など) に基づく Element の階層と、例えば行為主体の役割に基づく階層 (creator > author > screenwriter など) とが組み合わせられて形成される。この結果、場合によっては階層木の内外で同一 Element の複数回出現が発生し、階層木を巨大化させている。

マッピングに出現した範囲の RDA Element について、階層関係に基づき形成した階層木は、LRM プロパティへのマッピングでは 149、DC タームへのマッピングでは 156、MARC フィールドへのマッピングでは 582 であった。実際には階層を形成せず孤立した Element の数も、ここには含まれている。また、個々の階層木を構成するマッピング行数 (孤立した Element の行を含む) は、同一 RDA Element すなわち同一マッピングペアの複数回出現の結果、LRM へのマッピングは 43,909 行、DC タームへのマッピングは 3,324 行、MARC フィールドへのマッピングは 33,316 行にまで膨れ上がっている (表 2)。DC タームへのマッピング行数が他に比べて少ないのは、採用された RDA Element の階層上の位置づけが中位以下のものに限られていたことを表している。

1) LRM/DC へのマッピング : LRM プロパティについては IFLA が、DC タームについては DCMi が、それぞれ `rdfs:subPropertyOf` によって階層関係を定義している。なお、RDA マッピングで取り上げる範囲では、DC タームは多重階層ではなく、単純な階層木の構造である。

そこで、上記の RDA Element の階層木に従ってマッピングペアを配置したときに、マッ

ピング先の LRM プロパティまたは DC タームが、直上位 Element のマッピング先と同一あるいは階層関係として整合するかを照合した。照合結果は表 2 に示した通り、個別マッピング行に対して「最上位 Element」、「階層的に整合」、「階層的に不整合」に該当するかに分けた。なお、「部分的に整合」(一部のマッピング先は整合し、残りのマッピング先が不整合) に該当する例は、ここでは発生していない。

まず、指標 1-1 (RDA Element 階層と整合したマッピング先の割合) を求めると、「最上位 Element」と「階層的に整合」が占める割合となり、LRM では指標値 0.980、DC では 0.984 が得られた。同様に、指標 1-2 (マッピング先要素の照合を、直上位のものから、階層木の最上位 Element のマッピング先要素まで拡張した場合の整合性割合) は、LRM は 0.998、DC は 0.986 にそれぞれ上昇した。

また、同一 RDA Element の複数回出現が見られた場合、当該マッピング行においては「階層的に不整合」であっても、他のマッピング行では「階層的に整合」となる Element について、便宜的に当該マッピング行においても「階層的に整合」と見なした場合を指標 1-3 とした。この指標値では LRM は 0.990、DC は 0.988 との値が得られた。以上のように、RDA が概念モデルとして依拠する LRM へのマッピングにおいても、完全に整合的とはならない結果であった。

2) MARC へのマッピング : MARC フィールド自体には階層関係はなく、それに代えて包含関係を一部適用できる。a) インディケータの値「1」や「#」は、値「*」に合致し包含されるものと見なす。これにより、「B264 *1 \$a」は「B264 ** \$a」に包含される。b) サブフィールドが複数指定されているものとその部分集合は包含関係にあると見なす。例えば、「A510 ** \$a, b, c, d, g」に「A510 ** \$a, e」は包含されると見なす。さらに、単一 RDA Element から複数の MARC フィールドへのマッピングが設定されているため、その組み合わせにおける包含関係が発生する。これらの適用により、MARC フィールドへのマッピングにおいても階層整合性の照合が可能になり、ここでは「部分的に整合」に該当する事例も発生する。「最上位 Element」と「階層的に整合」が占める割合を表す指標 1-1 の値は 0.922、指標 1-2 が 0.957、指標 1-3 が 0.976 となった。

LRM/DC/MARC21 へのマッピングにおいては、いずれも RDA Element 階層とマッピ

ング先要素の階層とは高い整合性を見せているとはいえ、逸脱する事例が一定数見られる結果となった。

4. RDA エlement階層木ベースの指標

指標群 2 は、RDA エlement階層木単位の指標とし、マッピング先の共有に基づく書誌結合による RDA グラフ分割との類似度を表す指標群とする。RDA エlementは多重階層であるため正確にはグラフ構造であるが、ここでは指標群 1 において形成した階層木をそのままグラフ分割とする。階層木内での同一Elementの複数回出現は単一化されるが、複数の階層木にまたがって出現するものはそのまま複数出現となる（ただし、その数は限定的）。なお、マッピング先の共有に基づく書誌結合とは、RDA エlementとマッピング先からなる 2 部グラフから、RDA エlementへのプロジェクションをとりグラフ分割を行うことに等しい。

指標群は、a) RDA エlement階層木を正解と見なした場合の、書誌結合による RDA グラフ分割結果の精度・再現率・F 値を求めるもの、b) 正解をもたないグラフ分割結果の類似性を求める指標 Omega Index と Mean F1 score (MF1 ; 平均 F 値) とした。なお、a)は RDA エlementの「ペアベース」による計算であり、RDA 階層木をまたがった同一Elementの複数出現は障壁とはならない。b)の指標も、こうした複数出現、すなわちオーバーラップクラスに適合したものを採用した。Omega Index は Rand Index を拡張した指標であり、0 (完全不一致) から 1 (完全一致) の値をとる。MF1 は 0~1.0 の値をとる。

1) LRM へのマッピング : RDA エlement階層木による分割 (サブグラフ数 149) に対して、a) 書誌結合による分割 (同 103) は、その精度 0.979、再現率 0.084、F 値 0.155 という、再現率とそれに伴い F 値が低い結果になった (表 3)。次に、b) マッピング先 LRM プロパティ

の階層性を組み入れた上で書誌結合による分割を行った場合 (同 36) には、それぞれの値が 0.990、0.982、0.986 に上昇し、再現率の大幅な改善が見られた。同様に、Omega Index と MF1 の値も b)において大きく改善している。

2) DC へのマッピング : RDA エlement階層木による分割 (サブグラフ数 156) と、a) 書誌結合による分割 (同 32) との精度・再現率・F 値に対して、b) マッピング先 DC タームの階層性を組み入れた上で書誌結合による分割を行った場合 (同 22) には、精度とそれに伴う F 値が若干低下した。加えて、c) RDA エlement階層木の最上位Elementについて、マッピングには出現していない、さらに上位レベルのElementを調整して採用することによって階層木を統合した場合 (同 96)、書誌結合による分割結果との精度・F 値、そして Omega Index と MF1 の値は向上した。

3) MARC へのマッピング : RDA エlement階層木による分割 (サブグラフ数 582) と、a) 書誌結合による分割 (同 165) との類似度、b) マッピング先 MARC フィールドの階層的包含関係を組み入れた上で書誌結合による分割を行った場合 (同 150) との類似度、c) RDA エlement階層木のさらに上位レベルElementによって階層木を統合した場合 (同 165) との類似度は、いずれも同程度であった。加えて、d) マッピング先を MARC フィールド組み合わせの単位とした場合の書誌結合 (同 1,145) についても実験し、精度および F 値、そして Omega Index が向上する結果を得た。

引用文献

- 1) RDA Registry. <https://www.rdaregistry.info/>
- 2) 谷口祥一. RDA から MARC21 への複雑なマッピングを読み解く : 効果的な表示法の検討とネットワーク分析の適用. 第 72 回日本図書館情報学会研究大会発表論文集. 2024, p.45-48.

表 1 マッピングの基本集計結果 (値はいずれも異なり数)

マッピング元		マッピング先		マッピングペア	
LRM プロパティへのマッピング					
RDA エlement	3,024	LRM プロパティ	103	RDA エlement-LRM プロパティのペア	3,024
DC タームへのマッピング					
RDA エlement	1,153	DC ターム	33	RDA エlement-DC タームのペア	1,155
		DC ターム組み合わせ	34		
MARC フィールドへのマッピング					
RDA エlement	8,258	MARC フィールド	1,259	RDA エlement-MARC フィールドのペア	34,228
		MARC フィールド組み合わせ	1,070		

表2 RDA エlement階層との整合性指標群

	最上位エレ メント	階層的に 整合	部分的に 整合	階層的に 不整合	計
LRM プロパティへのマッピング					
指標 1-1: RDA エlement階層とマッピング 先 LRM プロパティ階層の整合	149 (0.3%)	42,870 (97.6%)	0 (0.0%)	890 (2.0%)	43,909 (100%)
指標 1-2: マッピング先 LRM プロパティ照合 を直上位から最上位までに拡張した場合	149 (0.3%)	43,693 (99.5%)	0	67 (0.2%)	43,909 (100%)
指標 1-3: 同一 RDA エlementの複数回出現 による整合箇所を考慮した場合	149 (0.3%)	43,325 (98.7%)	0	435 (1.0%)	43,909 (100%)
DC タームへのマッピング					
指標 1-1: RDA エlement階層とマッピング 先 DC ターム階層の整合	156 (4.7%)	3,114 (93.7%)	0 (0.0%)	54 (1.6%)	3,324 (100%)
指標 1-2: マッピング先 DC ターム照合を直上 位から最上位までに拡張した場合	156 (4.7%)	3,121 (93.9%)	0	47 (1.4%)	3,324 (100%)
指標 1-3: 同一 RDA エlementの複数回出現 による整合箇所を考慮した場合	156 (4.7%)	3,127 (94.1%)	0	41 (1.2%)	3,324 (100%)
MARC フィールドへのマッピング					
指標 1-1: RDA エlement階層とマッピング 先 MARC フィールドの階層的包含の整合	582 (1.7%)	30,142 (90.5%)	2,502 (7.5%)	90 (0.3%)	33,316 (100%)
指標 1-2: マッピング先 MARC フィールド照 合を直上位から最上位までに拡張した場合	582 (1.7%)	31,303 (94.0%)	1,359 (4.1%)	72 (0.2%)	33,316 (100%)
指標 1-3: 同一 RDA エlementの複数回出現 による整合箇所を考慮した場合	582 (1.7%)	31,938 (95.9%)	727 (2.2%)	69 (0.2%)	33,316 (100%)

表3 RDA エlement階層木ベースの指標群 (グラフ分割の類似度)

	精度(ペア ベース)	再現率(ペ アベース)	F 値(ペア ベース)	Omega Index	Mean F1 score
LRM プロパティへのマッピング					
a) RDA エlement階層木による分割(サブグ ラフ数 149) - 書誌結合による分割(103)	.979	.084	.155	.029	.238
b) RDA 階層木分割(149) - 書誌結合による分 割かつ LRM プロパティの階層性組み入れ(36)	.990	.982	.986	.923	.412
DC タームへのマッピング					
a) RDA エlement階層木による分割(156) - 書誌結合による分割(32)	.701	.968	.813	.783	.404
b) RDA 階層木分割(156) - 書誌結合による分 割かつ DC タームの階層性組み入れ(22)	.646	.988	.781	.744	.425
c) RDA 階層木かつ階層レベル調整による分割 (96) - 書誌結合による分割(32)	.752	.908	.822	.791	.441
MARC フィールドへのマッピング					
a) RDA エlement階層木による分割(582) - 書誌結合による分割(165)	.071	.999	.132	.093	.479
b) RDA 階層木分割(582) - 書誌結合による分 割かつ MARC フィールドの階層的包含の組み 入れ(150)	.070	.999	.132	.091	.500
c) RDA 階層木かつ階層レベル調整による分割 (530) - 書誌結合による分割(165)	.071	.998	.133	.093	.472
d) RDA 階層木分割(582) - 書誌結合による分 割かつマッピング先は MARC フィールド組み 合わせ(1,145)	.471	.229	.308	.297	.432