

Title	RDAからIFLA LRM/DC/MARC21へのマッピングとその整合性指標
Sub Title	Mappings from RDA to IFLA LRM, DC, and MARC 21 : indicators of mapping consistency
Author	谷口, 祥一(Taniguchi, Shoichi)
Publisher	三田図書館・情報学会
Publication year	2025
Jtitle	Library and information science No.94 (2025.) ,p.1- 23
JaLC DOI	
Abstract	【目的】 RDAからIFLA LRM, DCメタデータ語彙, MARC21への大規模かつ複雑なマッピングに対して, それを複雑にしている主たる要因とその状況を示し, さらにマッピングの整合性を適切に表す指標群を導入し計算することを目的とする。これらによって, マッピングの全体像理解に資することを意図している。 【方法】 RDA語彙v5.1.0において示された, 3つの対象スキーマそれぞれへのマッピングについて基本的な集計を行った。次に, マッピング元のRDAエレメント階層構造(ただし, 多重階層)とマッピング先要素の階層関係・包含関係が整合的となるかを, マッピングの整合性の要因として着目した。これらの状況を適切に表し評価するために, a) RDAエレメントベースの指標群と, b) RDAエレメント階層木ベースの指標群の2種類を導入した。 【結果】 基本的集計によって, 3つのマッピングの規模を表すRDAエレメント数, マッピングペア数はかなり異なる点を示した。次に, a) 3つのマッピングすべてにおいて, RDAエレメントとマッピング先要素の両者における階層関係は90%以上が整合していた。また, 整合性判定の条件を緩和することによって, 不整合となる事例を減少させることができたが, 依然として整合しない事例が一定数残された。b) 階層木によるRDAグラフ分割とマッピング先の書誌結合によるRDAグラフ分割とが類似する程度を複数の指標(エレメントペアベースの精度・再現率・F値, Omega Indexなど)によって示した。さらに, マッピング先の階層関係を組み入れた上での書誌結合, RDAエレメント階層木における最上位エレメントのレベル調整による階層木統合などを加えた場合について同様に指標値を求め, その効果を確認できた。これらの指標群によって, 3つのマッピングそれぞれの整合性とその相対的な位置関係を確認できた。 Purpose: This study investigates the complex mappings from RDA to IFLA LRM, Dublin Core (DC), and MARC 21. It aims to identify key sources of structural and semantic complexity and inconsistency within these mappings, and to introduce indicators that quantify their consistency. The goal is to support a clearer understanding of mapping structures across different metadata schemas. Methods: Using RDA Vocabularies v5.1.0, we analyzed the mappings to the three target schemas through basic statistical summaries. We focused on two primary sources of complexity: the polyhierarchical nature of the RDA elements and the alignment of hierarchical or containment relationships between the source and target elements. To evaluate these aspects, we developed two sets of indicators: (a) element-based metrics and (b) tree-based metrics derived from the hierarchical structure of the RDA elements. Results: The mappings differed significantly in size and structural characteristics. (a) In all three maps, over 90% of the RDA element-to-target pairs maintained their structural consistency. When the consistency criteria were relaxed, the number of mismatches decreased, although some inconsistencies persisted. (b) Comparisons between RDA graph partitioning based on hierarchical trees and that based on bibliographic coupling in the target mappings revealed a moderate degree of

	similarity, as measured by element pair-based precision, recall, F-measure, and the Omega Index. These measures improved when target-side hierarchical relationships were incorporated and the root levels of the RDA trees were adjusted. Collectively, these indicators offered a comparative perspective on structural alignment and consistency across the three mapping sets.
Notes	原著論文
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000094-0001

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

原著論文

RDA から IFLA LRM/DC/MARC21 へのマッピングとその整合性指標

Mappings from RDA to IFLA LRM, DC, and MARC 21: Indicators of Mapping Consistency

谷口 祥一

Shoichi TANIGUCHI

Résumé

Purpose: This study investigates the complex mappings from RDA to IFLA LRM, Dublin Core (DC), and MARC 21. It aims to identify key sources of structural and semantic complexity and inconsistency within these mappings, and to introduce indicators that quantify their consistency. The goal is to support a clearer understanding of mapping structures across different metadata schemas.

Methods: Using RDA Vocabularies v5.1.0, we analyzed the mappings to the three target schemas through basic statistical summaries. We focused on two primary sources of complexity: the poly-hierarchical nature of the RDA elements and the alignment of hierarchical or containment relationships between the source and target elements. To evaluate these aspects, we developed two sets of indicators: (a) element-based metrics and (b) tree-based metrics derived from the hierarchical structure of the RDA elements.

Results: The mappings differed significantly in size and structural characteristics. (a) In all three maps, over 90% of the RDA element-to-target pairs maintained their structural consistency. When the consistency criteria were relaxed, the number of mismatches decreased, although some inconsistencies persisted. (b) Comparisons between RDA graph partitioning based on hierarchical trees and that based on bibliographic coupling in the target mappings revealed a moderate degree of similarity, as measured by element pair-based precision, recall, F-measure, and the Omega Index. These measures improved when target-side hierarchical relationships were incorporated and the root levels of the RDA trees were adjusted. Collectively, these indicators offered a comparative perspective on structural alignment and consistency across the three mapping sets.

谷口祥一：元・慶應義塾大学文学部，〒108-8345 東京都港区三田 2-15-45

Shoichi TANIGUCHI: Formerly, Faculty of Letters, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan

e-mail: s.taniguchi@keio.jp  <https://orcid.org/0000-0002-6103-7902>

受付日：2025 年 6 月 16 日 改訂稿受付日：2025 年 8 月 28 日 受理日：2025 年 9 月 11 日

発行日：2025 年 12 月 31 日

DOI: 10.46895/lis.94.1

- I. はじめに
- II. RDA からのマッピングの基本的な特徴と集計
 - A. RDA エlement
 - B. IFLA LRM プロパティへのマッピング
 - C. DC タームへのマッピング
 - D. MARC フィールドへのマッピング
- III. RDA エlementベースの指標
 - A. LRM プロパティへのマッピング
 - B. DC タームへのマッピング
 - C. MARC フィールドへのマッピング
- IV. RDA エlement階層木ベースの指標
 - A. エlement階層木ベースの指標の概要
 - B. LRM プロパティと DC タームへのマッピング
 - C. MARC フィールドへのマッピング
- V. おわりに

I. はじめに

RDA 運営委員会は、最新の RDA 語彙から複数のメタデータ語彙やスキーマへのマッピングを策定し、RDA Registry 上で公開している (RDA Steering Committee, 2020-)。RDA エlementの主たるマッピング先には、a) IFLA LRM (図書館参照モデル) の属性・関連、b) DC メタデータ語彙のターム (DCMI Metadata Terms)、c) MARC21 書誌・典拠データフォーマットなどがある。これらのマッピングはいずれも人手により入念に作成されたものとはいえ、全体の規模が大きく、かつ複雑であり、全体的な把握や理解は困難と言わざるをえない。3R プロジェクト後の RDA エlement数の膨大さ、依拠する概念モデルの相違などを含むマッピング先語彙やスキーマとの構造的相違等が複合してか、マッピング全体の複雑さの程度は極めて高いと言える。

これらのマッピングに対して、策定方針などは公開されておらず、理解を助けるものがない。また、メタデータ変換ツールとして実装されているわけではない。3R プロジェクト以前の RDA から MARC21 へのマッピングは比較的シンプルであり (RDA Steering Committee, 2017a, 2017b)、RDA メタデータエディタ RIMMF パー

ジョン3は、このマッピングに基づきメタデータの変換機能を実装していた (RIMMF Project, 2022-)。そこではElement数の点でも、マッピング先 MARC フィールドの設定でも、限定的なものであった。それに対して、3R プロジェクト後の RDA (「オフィシャル RDA」とも呼ばれる) の語彙に対応した RIMMF バージョン4以降は、MARC レコードの読み込みなど、マッピングに基づく変換機能の実装は予定されていない (RIMMF Project, 2024-)。

マッピングは RDA エlementから他の語彙やスキーマのElement (またはそれに相当するもの) への対応関係を意味するが、RDA エlement側の関係や制約がマッピング先でも体系的に保たれているという状態を、ここではマッピングの「整合性」と呼ぶ。例えば、RDA エlement間の階層関係がマッピング先でもそれが保持されるべきであるが、必ずしもそうではない事例もある。同様に、RDA エlement階層木単位で見たときに、マッピング先集合が相互に排他的あるいは包含の関係などになるべきであるが、部分的に重なるなど、必ずしもそうではない状況が生じうる。

本研究の目的は、公表されている RDA から IFLA LRM, DC メタデータ語彙, MARC21 へのマッピングについて、a) 基本的な集計等に

よってその概要を提示し、併せてb)どのような点において複雑であるのかを示すことである。加えて、c)その整合性（換言すれば複雑性）を適切に表す指標（尺度）を導入し、実際に指標値を求めることを目的とする。後述する通り、RDAエレメントは階層性をもち、しかも上位エレメントを複数もつという多重階層として定義されている。b)は例えばこうした階層性に起因してマッピングがどのような複雑さを作り出しているのかを問うこと、c)はマッピングの整合性を適切に表す指標にはどのようなものがあるのかを問い、それらを具体的に計算して求めることを指している。これらによって最終的には、マッピングの全体像理解に資することを意図している。

筆者は、RDAからのマッピングのうち、DCメタデータ語彙とMARC21へのマッピングを取り上げ、その基本的集計、RDAエレメントの階層構造に依拠した表示法およびマッピング先の整合性確認などを試行し、その成果を既に報告している（Taniguchi, 2025）。本研究はその続編に当たるが、a)IFLA LRMを新たに追加し3つのマッピングを取り上げ、それぞれのマッピングの全体像を改めて提示し相互比較を行う点、さらにb)RDAエレメント階層構造に依拠したマッピング先の整合性を含め、マッピングの整合性を表す指標群を複数提案し、それらの値を求める点で、前稿と異なる。

3Rプロジェクト後のRDAによるマッピングを直接取り上げた研究は見当たらない。関連するプロジェクトとしては米国ワシントン大学図書館によるものがあり、そこではRDA運営委員会が策定していない、MARC21フォーマットから最新RDAへのマッピングの策定と変換ツールの開発が試行されている（University of Washington Libraries, 2021-）。ただし、書誌データフォーマットのみ対象としており、典拠データは対象外である。RIMMFバージョン4以降における、変換機能の欠落を補うことを目的としている。

他方、RDAエレメント以外の多様なメタデータのマッピング、アラインメント、そしてクロスウォーク（crosswalk）に目を向けると、研究

やプロジェクトは極めて数が多い。これらはメタデータの意味的相互運用性（semantic interoperability）の問題として扱われることも多い。このようなトピックを取り上げたものに、ZengとQin（2022）による包括的な教科書、そして複数のレビューがある（Gaitanou et al., 2024; Haslhofer, 2008; Woodley, 2016）。さらには、メタデータのデータ要素を含めてメタデータ語彙は、オントロジーの一種と見ることができ、そこにはオントロジーのマッピングやアラインメントを扱った膨大な研究が存在する。これらの領域と成果についての包括的な解説書にはEuzenatとShvaiko（2013）によるものがある。また、包括的なレビュー論文も複数ある（Babalou et al., 2016; Osman, 2021; Otero-Cerdeira, 2015; Scharffe et al., 2014; Thiéblin, 2019）。

これらの中にはマッピングやアラインメントに関わる「不整合」（mismatches）について区分を示した研究もある。例えばScharffeら（2014）は、大きくConceptualization, Relation, Model coverage and granularity, Explication, Encodingの不整合に区分し、それらをさらに細分し、Relationの不整合は、Structure, Attribute assignment, Attribute-typeに分けている。この「Structure」不整合について、“対応関係の定義域と値域は、互いに整合するよう制約されるべきである”（Scharffe et al., 2014, p.17）としており、本研究で取り上げる問題は主にここに属することになる。ただし、本研究に類似する研究は見当たらない。

II. RDAからのマッピングの基本的な特徴と集計

A. RDAエレメント

2024年7月時点のRDA語彙v5.1.0¹⁾とそのマッピングを使用した。その後も語彙は随時更新されているが、現時点（2025年6月）のバージョン番号からは大幅な変更はなされていないものと仮定できる。RDA語彙からのマッピングは、複数の語彙やスキーマが対象とされており、本研究はそのうち、a)IFLA LRMの属性・関連、b)DCメタデータ語彙のターム、c)MARC21書誌・典

拠データフォーマットへのマッピングを取り上げる。これらのマッピングの特徴は、a) 出現する RDA エlement 数の膨大さに加えて、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) マッピング先要素はスキーマに依存し 1 または多となる、さらに d) 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。

先ずマッピング元の RDA であるが、実体および値の語彙は RDF クラス、Element は RDF プロパティとして定義され、マッピングはこれらを用いて示されている。推奨されている QName を用いて表すと、例えば体現形のElement 「title proper」(本タイトル) のURI は rdam:P30156, 「title of manifestation」(体現形のタイトル) は rdam:P30134 と表される。接頭辞「rdam:」は名前空間「<http://rdaregistry.info/Elements/m/>」を表しており、定義域となる実体ごとに接頭辞すなわち名前空間が異なる。これら2つのプロパティは、定義域としてクラス rdac:C10007 (体現形) を取り、値域の指定はない。なお、同じElement に対して、値域をURI リソースに限定したプロパティ <http://rdaregistry.info/Elements/m/object/P30156>, 値域をリテラルに限定したプロパティ <http://rdaregistry.info/Elements/m/datatype/P30156> を別途定義しているが、これらはマッピングには用いられていない。

RDA Element は、現在 3,000 以上あり、追加・削除・修正が継続して行われている。なお、本研究ではマッピングに登場した範囲のElement のみが対象となり、マッピング先のスキーマによってその範囲は異なる。

B. IFLA LRM プロパティへのマッピング

3R プロジェクト後の RDA は、概念モデルである IFLA LRM に準拠しており、この依拠する概念モデルに対するマッピングをも示している。LRM は実体関連モデルを用いて構築されており、同時にリンクトデータとして参照できるよう IFLA (2020) によって実体は RDF クラス、属性と関連は RDF プロパティとして定義されて

いる。このうち、RDA Element からのマッピングに登場するのが、LRM の属性と関連であるプロパティ群となる。例えば、LRM が定義する属性「has category of res」(res のカテゴリ) はプロパティ lrmer:E1A1, 関連「has appellation」(呼称をもつ) は lrmer:R13 としている。接頭辞「lrmer:」は名前空間「<http://iflastandards.info/ns/lrm/lrmer/>」を表す。なお、LRM が定義する属性と関連は、「代表的」すなわち粒度の大きなものにとどめられている。

RDA Element からのマッピングもこれらプロパティを用いて表されており、例えば「rdam:P30156 rdfs:subPropertyOf lrmer:R13.」, 「rdam:P30134 rdfs:subPropertyOf lrmer:R13.」としている。rdam:P30156 と rdam:P30134 はともに LRM プロパティ lrmer:R13 のサブプロパティの関係にある (rdfs:subPropertyOf) として、マッピングを表している。すなわち、マッピング元プロパティを、マッピング先プロパティのサブプロパティ (下位プロパティ) に位置づけることによってマッピングとしている。

この点に依拠すれば、マッピング元と先の定義域および値域は相互に矛盾しない関係にあることになる。しかし、lrmer:R13 は定義域 lrmer:E1 (Res), 値域 lrmer:E9 (Nomen) と定義されている。上位プロパティに位置づけられた lrmer:R13 が値域 lrmer:E9 を指定しているのに対して、サブプロパティとした rdam:P30156 や rdam:P30134 は前述の通り、値域を指定していない。RDA Element が取る値は URI リソースの場合と単なるリテラルの場合との両方があり得るため、整合しない。しかし、示されたマッピングは主に意味的な対応を表すことを主眼としており、リンクトデータとしての論理的整合性を保つことを重視していないと捉えるべきである。一方、定義域としたクラス rdac:C10007 も lrmer:E1 のサブクラスでなければならないことになるが、この条件は満たされていると見なせる。

本研究で採用したバージョンのマッピングは、RDA Element 数 (異なり数) 3,024, マッピング先の LRM プロパティ数 (異なり数) 103 と

第1表 マッピングの基本集計結果

マッピング元とマッピング先		マッピングペアとマッピング元と先の平均	
<u>LRM プロパティへのマッピング</u>			
RDA エlement	3,024	RDA エlement-LRM プロパティのペア	3,024
LRM プロパティ	103	[平均 1.0 (SD 0.0) - 平均 29.4 (SD 74.0)]	
<u>DC タームへのマッピング</u>			
RDA エlement	1,153	RDA エlement-DC タームのペア	1,155
DC ターム	33	[平均 1.0 (SD 0.1) - 平均 35.0 (SD 74.4)]	
DC ターム組み合わせ	34		
<u>MARC フィールドへのマッピング</u>			
RDA エlement	8,258	RDA エlement-MARC フィールドのペア	34,228
MARC フィールド	1,259	[平均 4.1 (SD 2.8) - 平均 27.2 (SD 111.7)]	
MARC フィールド組み合わせ	1,070		

なり、これらを組み合わせたマッピングペアは3,024であった（第1表）。RDA エlementごとに単一 LRM プロパティへのマッピングしている（マッピング先の最大数1）。

C. DC タームへのマッピング

汎用的なメタデータ語彙の代表と目されるダブリンコアには、2種類の語彙が定義されている。DCMES (DC Metadata Element Set) と呼ばれる基本15要素のみからなる語彙と、より拡張的なDCメタデータ語彙 (DCMI Metadata Terms) とがある (DCMI Usage Board, 2020, January 20)。RDA エlementからのマッピングに登場するのは後者であり、そこでのプロパティがマッピング先要素となっている。dct:title, dct:creator などがあり、接頭辞「dct:」は名前空間「<http://purl.org/dc/terms/>」を表す。これらプロパティを、以後、「DC ターム」と呼ぶ。

RDA エlementからDCタームへのマッピングについても、「rdam:P30156 rdfs:subPropertyOf dct:title.」, 「rdam:P30134 rdfs:subPropertyOf dct:title.」のように、RDA エlementをDCタームのサブプロパティと位置づけている。ただし、この場合、dct:titleの値域はリテラルであるため、厳密に捉えればLRMプロパティの場合と同様、リンクトデータとしての論理的整合性には問題が残る。

マッピングに登場したRDA Element数1,153, DCターム数33であり、マッピングペア数

1,155であった（第1表）。マッピング先DCタームが1つのものが1,151 Element, 2つのものが2 Elementであった（マッピング先の最大数2）。後者に該当するのは、rdac:P20069 (summarization of content) と rdaw:P10109 (summary) であり、いずれもマッピング先はdct:abstractとdct:descriptionの組み合わせであった。

D. MARC フィールドへのマッピング

MARC21 書誌データフォーマットと典拠データ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へのマッピングを含めたものとする。そのマッピングとは、例えばRDA Element rdam:P30156 については、「rdam:P30156 rdakit:hasM21 "MARC 21 Bibliographic 245**\$a [unstructured description]".」とし、MARC21 書誌データ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 (第1・2インディケータの「*」は任意の値を表す), 「unstructured description」(非構造記述) の値としてマッピングされるとしている。同様に、RDA Element rdam:P30134 については、「rdam:P30134 rdakit:hasM21 "MARC 21 Bibliographic 245**\$a [unstructured description]".」, "MARC 21 Bibliographic 500**\$a [unstructured description]".」と指定している。先ほどのマッピング先に加えて、フィールド500 (General note), サブフィールドa (General note) にも対

応づけている。これは、指定された2つのマッピング先要素に必ず変換され記録されることを指示したわけではなく、いずれか1つへの記録も許容されると見るべきである。

なお、マッピングを指示するプロパティ `rdakit:hasM21` は RDA Registry が導入したものであり、マッピング先を表す目的語がリテラルで示されている。RDA が定めるエレメントの規定では、「unstructured description」以外に、「structured description」(構造記述)、「identifier」(識別子)、「IRI」(URI と基本的に同義)という、4つの「recording method」(記録方法)が適用されており、単一エレメントがその記録方法に依存して、異なるマッピング先が指定されている。ただし、すべてのエレメントにおいて4つすべての記録方法が適用されるわけではない。

以下、MARC21 データフォーマット (Library of Congress, 1999-) へのマッピングに対する本研究での扱いをまとめる。

- 1) 示されたマッピングではマッピング先が MARC フィールド等と記録方法との組み合わせとしているが、本研究ではマッピング元を「エレメント+記録方法」の単位とし、マッピング先は記録方法を含まない、MARC フィールド等のみとする。つまり、RDA による「記録方法」をマッピング元に移行し、エレメントと記録方法が組み合わせられ、それに従い形成された値がマッピング先としてどの MARC フィールド等に記録されるべきかと捉えることが、実際のメタデータマッピングに即していると考えた。同一エレメントであっても、適用する記録方法すなわち値の形式に依存して、マッピング先が変更される事態を素直に表現できるものとした。本研究ではリンクトデータを前提とした議論を展開するものではないが、敢えて RDF による表現を取るならば、下記のように表すことができる(それぞれの URI リソースは仮のもの)。

```
rdam:P30156_unstructured_description
rdakit:hasM21 m21:B245_**_$a.
```

本研究では以後、MARC へのマッピングにおいては、マッピング元の「RDA エレメント+記録方法」を便宜的に「RDA エレメント」と呼ぶ。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。これらは MARC 固定長フィールドの場合も含め、単に簡便な呼称とするためである。併せて、書誌データフォーマットの場合は接頭辞「B」、典拠データフォーマットの場合には「A」を付けて区別する。エレメントに組み入れた記録方法の表記も、本文中では「unstructured」・「structured」のように略記する。

- 2) マッピング先 MARC フィールドには、「B245**\$a, c, p」や「B710**\$a, b, c, d, f, g, k, l, n, p, t, u (if subfield \$t is used)」といった指定も登場するが、指定されたままで単一マッピング先としてそれぞれを扱う。固定長フィールドについても、「B007/01」などから、「B006/01-02 (continuing resources)」, 「B008/07-10 (if 008/06=q, s or t)」などまで、多様な指定が登場するが、それぞれを単一のマッピング先としてそのまま扱う。
- 3) マッピング先として、「B245**\$7 (dpecou)」や「B245**\$7 (dpermw) (data provenance)」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ものは、本研究の対象外としてすべて除外した。「B245**\$7」(サブフィールド7は「Data provenance」)には、サブフィールドaなどに記録したタイトル等に関わり、その context of use (rdan:P80067), note on metadata work (rdaw:P10402), source consulted (rdaw:P10406) など、値の由来に該当する RDA エレメント群からのマッピングが集められている。このようなマッピングを含めて検討することは、議論の複雑さを増大させ、見通しを一層悪くさせるため、本研究では一律に除外した。

上記の前処理を加えた後のマッピング集計結果

は、以下の通りである。

- 1) 出現した RDA エLEMENT の異なり数は 8,258 (記録方法を組み合わせない、ELEMENT のみの場合は 2,481)、出現した MARC フィールド異なり数は 1,259 であった (第 1 表)。RDA ELEMENT から個別の MARC フィールドへのマッピングペアは合計 34,228 となり、そのうち書誌データフィールドへのマッピングは 19,104、典拠データフィールドへのそれは 15,124 であった。
- 2) RDA ELEMENT ごとに平均 4.1 マッピングペア (SD 2.8)、すなわち 4.1 個のマッピング先 MARC フィールドが指定されていた。最大は rdae:P20298 [identifier] (related work of expression) の 30 ペアであった。他方、単一マッピング先のみもつ RDA ELEMENT は 1,128 (全体の 13.7%) であった。

単一 RDA ELEMENT から複数 MARC フィールドへのマッピングは、マッピング先が「論理和」の場合と「排他的論理和」の場合とが混在している。前者は単一の値をマッピング先の複数フィールドに並行して記録することを許容する場合であり、後者は示されたフィールドのいずれか 1 つに事例に応じて記録する扱いに該当する。いずれにしても、可能性のある MARC フィールドを幅広く網羅したマッピングといえよう。

III. RDA ELEMENT ベースの指標

A. LRM プロパティへのマッピング

1. RDA ELEMENT 階層と LRM プロパティ階層
RDA からのマッピングの整合性を表す指標の 1 つとして、RDA ELEMENT ごとにその階層に整合したマッピング先を指定しているかという、ELEMENT ベースでの割合を表す指標を採用する。

RDA ELEMENT は階層構造をもち、以前は「ELEMENT・サブタイプ」と名付けられていた関係であり、rdfs:subPropertyOf によって定義される。例えば、先に説明例に用いた rdam:P30156 と rdam:P30134 には、階層関係「rdam:P30156 rdfs:subPropertyOf rdam:P30134」が定義されている。なお、以前に存在した「サブELEMENT」

という関係は、RDA は現在採用していない。

さらに、大きな特徴として、単一ELEMENT が複数の上位ELEMENT を持ち得る「多重階層」(poly-hierarchy) として定義されている。この多重階層は複数の要因によってもたらされ、例えば定義域とするクラスの階層 (RDA entity>agent>collective agent>corporate body など) に基づくELEMENT の階層と、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発生する。この結果、rdaw:P10436 (author person) の上位ELEMENT は rdaw:P10437 (creator person of work) と rdaw:P10061 (author agent) の 2 つとなる。さらに、rdaw:P10437 の上位は、rdaw:P10312 (related person of work) と rdaw:P10065 (creator agent of work) となり、他方 rdaw:P10061 の上位は rdaw:P10065 のみとなるなど、幾分複雑である。

マッピングに出現した範囲で RDA ELEMENT の直上位・直下位の関係でつながるものから階層木、すなわち最上位となるELEMENT から順次下降する木構造を形成することができる。中間のノードが出現していない離れた上位下位関係にあるものは採用していない。前述の多重階層による同一ELEMENT の複数回出現はそれぞれ個別の出現 (階層木における個別ノード) として扱った結果、階層木の内外で同一ELEMENT の複数回 (場合によっては多数回) 出現が発生し、結果的に階層木を巨大化させている (具体例は後述する)。

形成した階層木は、LRM プロパティへのマッピングでは 149 となった (第 2 表)。そのうち、複数階層からなる木が 28、階層を形成しない孤立ELEMENT が 121 であった (平均階層数 1.4)。個々の階層木を構成するノード数 (マッピング行数) は、同一ELEMENT の複数回出現によって、合計 43,909 にまで膨れ上がっている。マッピングペア数 3,024 から見ると、その 14 倍以上の数である。階層木ごとのノード数の平均は 294.7 (SD 3573.4)、平均ELEMENT 異なり数 20.3 (SD 224.3) であった。

最大規模の階層木は、rdax:P00028 (related entity of RDA entity) を最上位ELEMENT とす

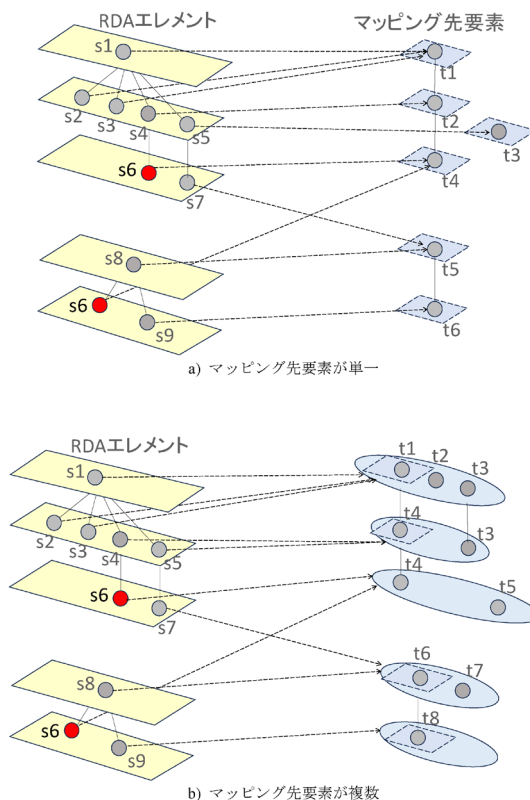
第2表 RDA エlement階層木の集計結果

	階層木	複数階層の木	単一階層の木	平均階層 (SD)	最大階層	マッピング行数
LRM プロパティへのマッピング	149	28	121	1.4 (1.1)	12	43,909
DC タームへのマッピング	156	56	100	1.7 (1.3)	8	3,324
MARC フィールドへのマッピング	582	228	354	2.0 (1.7)	9	33,316

る木であり、12階層、構成するノード数43,621、含まれるRDA Element異なり数2,739に及んでおり、この階層木1つがマッピング全体の大半を占める結果となった。最上位Element rdax:P00028の下に、第2階層として rdaw:P10405 (related entity of work), rdaa:P50478 (related entity of agent) など、個別実体に対応するElementと、加えて rdax:P00001 (related RDA entity of RDA entity) が位置づけられている。また、第3階層には、例えば rdaw:P10307 (related RDA entity of work) が上記の rdaw:P10405 と rdax:P00001 の両者の下に出現し、rdaa:P50304 (related RDA entity of agent) が上記の rdaa:P50478 と rdax:P00001 の両者の下に出現するといった複雑な構図を呈している。なお、2つ以上の階層木にまたがって出現しているElementは、rdam:P30152 (parallel other title information of series) など、計3つ見られた。

一方のLRM プロパティについては、IFLA が rdfs:subPropertyOf によって、それぞれ階層関係を定義している (IFLA, 2020)。実体の属性に該当するプロパティのうち、lrmer:E2A1 (has category of work) や lrmer:E4A1 (has category of carrier) などが lrmer:E1A1 (has category of res) のサブプロパティとされており、これら以外は階層を形成していない。実体間の関連に該当するプロパティは、すべて lrmer:R1 (is associated with res) のサブプロパティとする、2階層の木を構成している。すなわち、LRM プロパティは多重階層ではない。

以上のように、RDA ElementおよびLRM プロパティの両者に階層関係が定義されている点に依拠して、RDA Elementの階層木に従ってマッピング行を配置したときに、マッピング先の



第1図 RDA マッピングの説明図

LRM プロパティが直上位Elementのマッピング先と同一、あるいは階層関係にあるかを照合することができる。単純なマッピングであるほど、こうした整合性が保たれているものと仮定できる。逆に、この点の整合性が低いほど、マッピングは理解しにくい複雑なものとなる。

説明用に作図した第1図-a) では、RDA Elementが2つの階層木を構成しており、それぞれの最上位Elementはs1とs8となる。s1の下位Elementがs2からs5、さらにs4の下位が

s6, s5の下位がs7となる。s6はs8をも上位エレメントとする多重階層の例であり、そのためs8を最上位とする階層木にも出現している。各エレメントからのマッピングは、s1からt1, s2からt1といったように点線を用いて示した。併せて、マッピング先要素にも階層関係があり、ここではt1>t2>t4とt5>t6に階層関係があるとした。こうした設定下では、マッピング元であるRDAエレメントの階層関係とマッピング先要素の階層関係との整合性の観点からは、s5からt3へのマッピングペア、s7からt5へのペアは、それぞれマッピング先要素が直上位エレメントのマッピング先と同一あるいは階層関係とはなっておらず、この点で不整合となる。s6からt4へのペアについては、s4を上位エレメントとする場面では階層的に整合し、s8を上位とする場面では不整合となる。

2. RDAエレメントベースの整合性指標とその値

上述の整合性を照合するため、第2図にその一部を示したマッピングデータテーブルを形成した。このデータ全体は、後述するDCタームおよびMARCフィールドへのマッピングデータテーブルとともに、別途公開している（本文末尾にURLを記載）。

同図の上半分がLRMプロパティへのマッピングに該当し、ここでは階層木の番号（「Group no.」）が5と27についてその一部を示した。後者の階層木は前述した最大規模の階層木であり、説明用にその一部を切り取って示した。RDA階層木における階層レベル（「Hierarchy level」）は、階層木の最上位が1、その下位が2という順に付与している。この階層レベルに合致するようRDAエレメントは字下げして配置した。階層木「5」の最上位エレメントrdae:P20217（capture information）は、LRMプロパティlrmer:E1A2（has note）にマッピングし、その下位エレメントrdae:P20004（date of capture）はlrmer:R35（has association with time-span）に、rdae:P20218（place of capture）はlrmer:R33（has association with place）にマッピングする

ことを表している。マッピング先のLRMプロパティは、同一のものが同じ列となるよう配置した。RDA階層木内でのマッピング先LRMプロパティの異なりを示す意図であり、LRMプロパティの階層関係を表してはいない。

加えて、第2図の中間の列には、当該エレメントの全出現回数（「Occurrence count」）、階層整合性コード（「Hierarchical consistency code」）のCode1, Code2, Code3を記録した。それぞれの意味とそれらに基づく整合性の指標は、以下の通りであり、併せて集計結果を第3表にまとめた。

1) 指標 1-1: RDAエレメント階層とマッピング先LRMプロパティ階層の整合性。図中のCode1は、当該マッピング行のLRMプロパティが直上位エレメントのマッピング先LRMプロパティと整合しているかを、コードをもって示した。

- ・Code1=0: 最上位のエレメント（孤立エレメントを含む）を表し、その数は149（全体の0.3%）で、階層木の数となる。この数は指標1-2, 1-3においても変更はない。

- ・Code1=1: 当該エレメントのマッピング先が、直上位エレメントのマッピング先と同一またはその下位にあり、階層的に整合することを表す。該当する行は42,870、全体の97.6%を占める。第2図の階層木「5」に示したrdae:P20309（note on capture）は直上位エレメントrdae:P20217のマッピング先と同一であり、これに該当する。同様に、階層木「27」のrdam:P30156やrdam:P30128（variant title of manifestation）などは、rdam:P30134と階層関係にあり、そのマッピング先は同じLRMプロパティlrmer:R13（has appellation）である。さらに上位のrdam:P30277（appellation of manifestation）のマッピング先は同じlrmer:R13であるが、その上位のrdam:P30274（related nomen of manifestation）になると、マッピング先はlrmer:R1（is associated with res）となる。しかし、lrmer:R13はlrmer:R1のサブプロパティであるため、ここでも階層的に整合となる。

- ・Code1=2: 直上位エレメントのマッピング先

Group no.	Hierarchy level	RDA element	Element name	Occurrence count	Hierarchical consistency code: Code1 Code2 Code3	LRM property/DC term
Mapping to LRM properties						
5	1	rdac:P20217	capture information	1	0 0	lrmer:E1A2 has note
5	2	rdac:P20004	date of capture	4	3 1 3	*lrmer:R35 has association with time-span
5	2	rdac:P20218	place of capture	4	3 1 3	*lrmer:R33 has association with place
5	2	rdac:P20309	note on capture	1	1	lrmer:E1A2 has note
27	1	rdac:P00028	related entity of RDA entity	1	0 0	lrmer:R1 is associated with res
27	2	rdam:P30330	related entity of manifestation	1	1 1	lrmer:R1 is associated with res
27	3	rdam:P30264	related RDA entity of manifestation	2	1 1	lrmer:R1 is associated with res
27	4	rdam:P30274	related nomen of manifestation	3	1 1	lrmer:R1 is associated with res
27	5	rdam:P30277	appellation of manifestation	4	1 1	lrmer:R13 has appellation
27	6	rdam:P30134	title of manifestation	5	1 1	lrmer:R13 has appellation
27	7	rdam:P30128	variant title of manifestation	6	1 1	lrmer:R13 has appellation
27	7	rdam:P30131	abbreviated title	5	1 1	lrmer:R13 has appellation
27	7	rdam:P30156	title proper	6	1 1	lrmer:R13 has appellation
27	8	rdam:P30203	parallel title proper	6	1 1	lrmer:R13 has appellation
27	9	rdam:P30204	parallel title of series	11	1 1	lrmer:R13 has appellation
27	7	rdam:P30157	title of series	5	1 1	lrmer:R13 has appellation
27	8	rdam:P30204	parallel title of series	11	1 1	lrmer:R13 has appellation
Mapping to DC terms						
23	1	rdam:P30134	title of manifestation	1	0 0	det:title
23	2	rdam:P30128	variant title of manifestation	1	1 1	det:title
23	2	rdam:P30131	abbreviated title	1	1 1	det:alternative
23	2	rdam:P30156	title proper	1	1 1	det:title
23	3	rdam:P30203	parallel title proper	1	1 1	det:title
23	2	rdam:P30157	title of series	1	1 1	det:title
39	1	rdam:P30327	contributor agent to aggregate	1	0 0	det:contributor
39	2	rdam:P30315	contributor agent of cartography	1	1 1	det:contributor
39	3	rdam:P30312	surveyor agent	1	3 3	*det:publisher
39	4	rdam:P30337	surveyor person	3	1 3	det:publisher
39	4	rdam:P30366	surveyor collective agent	3	1 3	det:publisher
39	5	rdam:P30395	surveyor corporate body	6	1 3	det:publisher
39	5	rdam:P30424	surveyor family	6	1 3	det:publisher
39	3	rdam:P30338	contributor person of cartography	2	1 1	det:contributor
39	4	rdam:P30337	surveyor person	3	3 1 3	*det:publisher

第2図 LRM プロパティ/DC タームへのマッピング階層表示

第3表 RDA エlementベースの指標値

	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	整合性 指標値
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%)	0.980
指標 1-2: 同一 RDA エlementの複数回出現 による整合箇所を考慮	149 (0.3%)	43,325 (98.7%)	0	435 (1.0%)	43,909 (100%)	0.990
指標 1-3: 最上位エレメントのマッピング先 LRM プロパティまで照合範囲を拡張	149 (0.3%)	43,693 (99.5%)	0	67 (0.2%)	43,909 (100%)	0.998
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%)	0.984
指標 1-2: 同一 RDA エlementの複数回出現 による整合箇所を考慮	156 (4.7%)	3,127 (94.1%)	0	41 (1.2%)	3,324 (100%)	0.988
指標 1-3: 最上位エレメントのマッピング先 DC タームまで照合範囲を拡張	156 (4.7%)	3,121 (93.9%)	0	47 (1.4%)	3,324 (100%)	0.986
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%)	0.922
指標 1-2: 同一 RDA エlementの複数回出現 による整合箇所を考慮	582 (1.7%)	31,938 (95.9%)	727 (2.2%)	69 (0.2%)	33,316 (100%)	0.976
指標 1-3: 最上位エレメントのマッピング先 MARC フィールドまで照合範囲を拡張	582 (1.7%)	31,303 (94.0%)	1,359 (4.1%)	72 (0.2%)	33,316 (100%)	0.957

に部分的に包含され、一部は逸脱するという、「部分的に整合」を表す。LRM へのマッピングにおいては該当例がない。

- ・Code1=3: 直上位エレメントのマッピング先から逸脱し、階層的に不整合となることを表す。該当事例数 890 (2.0%) であった。第2図の階層木「5」では、rdac:P20004 と rdac:P20218 の行がこれに該当し、それぞれのマッピング先 LRM プロパティは直上位エレメントのマッピング先と階層的な関係になく、整合していない。このような事例には、LRM プロパティに「*」を前置し明示した。

Code1 の値を集計し、RDA エlement階層と整合したマッピング先の割合、すなわち「最上位レベルのエレメント」(Code1=0)と「階層的に整合」(Code1=1)が全体に占める割合を指標化すると、値 0.980 が得られた。マッピング行全体の 2% を除いて階層関係に整合したマッピングとされていることが判明した。なお、上記の値を 1 から減じた値 0.020 をもって (すなわち不整合と

なる割合によって)、複雑さの程度を表す指標値とすることも可能である。

- 2) 指標 1-2: 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の整合性。同一 RDA エlementが複数回出現するとき、それぞれの出現箇所である行によって上記の Code1 の値が異なる場合がある。Code2 はそのような場合に、当該箇所における Code1 の値と異なる値が他の箇所で発生しているときに、その値を示した。

例えば、第2図の階層木「5」に出現した rdac:P20004 と rdac:P20218 は全体でそれぞれ 4 回ずつ出現しており (Occurrence count が「4」)、前述した通り、図示した箇所では「階層的に不整合」(Code1=3)である。しかしながら、Code2 の値は「1」と記しており、これは図示されていない他の箇所では Code1=1 (階層的に整合) が少なくとも 1 例以上あることを表している。このような事例を、便宜的にすべての箇所では階層的に整合と見なしたときの、整合する割合

を指標化したのが指標 1-2 である。455 箇所がこれに該当し、階層的に整合する事例に加わった結果、「最上位レベルのエレメント」と「階層的に整合」が全体に占める割合は、0.990 に増加した。

なお、出現回数が 2 以上で、かつ Code2 の値が空値の事例は、当該エレメントの Code1 の値がすべての出現箇所と同じであったことを示している。

3) 指標 1-3: 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の整合性。マッピング先要素の照合を直上位のものから、当該 RDA エレメント階層木の最上位エレメントのマッピング先要素まで拡張した場合の、整合性照合結果を Code3 とした。すなわち、Code3=1 (階層的に整合) は、最上位エレメントから直上位エレメントまでのマッピング先のいずれかに等しいか、またはその下位の場合とし、Code3=3 (階層的に不整合) は最上位エレメントから直上位エレメントまでのいずれのマッピング先からも逸脱する事例を表す。

第 2 図に示した範囲では Code3 の値が Code1 の値から変更になる事例はないが、LRM へのマッピング全体において計 823 箇所がこれに該当し、階層的に整合となる事例に移行している。その結果、Code3=3 は 67 事例にとどまった。これらゆえ、RDA エレメント階層と整合したマッピング先の割合として指標化すると、値 0.998 が得られた。

最大の階層木「27」を形成する最上位エレメント rdax:P00028 のマッピング先は lrmer:R1 である。また、LRM プロパティのうち、実体間の関連に該当するもの (「lrmer:R..」の形式の URI をとるもの) は、すべて lrmer:R1 のサブプロパティに位置づけられている。このため、先の最大階層木に含まれるエレメントで、かつ LRM の関連に該当するプロパティをマッピング先とする行は、すべて Code3=1 となり、整合的な事例に属することになる。

以上の通り、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は、当然ながら整合的となる部分が殆どを占めるとはいえ、すべ

てが整合的とはならない結果であった。また、指標 1-1 よりも、照合条件を緩和した 1-2 と 1-3 が (さらにはこの順で) 整合性指標値が若干上昇する結果を示した。

B. DC タームへのマッピング

DC タームへのマッピングに出現した範囲で RDA エレメントの階層関係に基づき階層木を形成したところ、156 の階層木に分かれた (第 2 表)。そのうち、複数階層による木が 56、階層を形成しない孤立エレメントが 100 であった (平均階層数 1.7)。階層木の平均エレメント異なり数 7.5 (SD 31.0)、エレメント出現回数 21.3 (SD 151.0) となった。

個々の階層木を構成するノード数 (孤立したエレメントの木を含む) は、同一エレメントの複数回出現を含めて 3,324 であった。最大の階層木は、rdax:P20053 (creator agent of expression) を最上位エレメントとする木であり、8 階層、構成するノード数 1,768、RDA エレメント異なり数 293 であった。rdax:P20502 (voice actor corporate body) など、5 つのエレメントは階層木内で計 21 回ずつ出現している。また、rdaw:P10100 (description of work) など、18 エレメントは複数の階層木にまたがり出現している。

ノード数 (マッピング行数) が LRM へのマッピングに比べて少ないのは、採用されている RDA エレメントが、定義された RDA エレメント群全体の階層上における位置が中位以下に限定されていることによる。具体的には、定義域を RDA entity とするプロパティ (接頭辞「rdax:」をもつもの) が、ここでは出現していない。

第 2 図の下半分に、DC タームへのマッピング事例について階層表示をもって示した。階層木「23」は最上位エレメント rdam:P30134 であり、その直下に rdam:P30128、rdam:P30156 などがある。rdam:P30134 は、LRM へのマッピングでは階層レベル 6 に配置されており、こうした点において大きく相違する。また、LRM へのマッピングでは登場した rdam:P30204 (parallel title of series) は採用されていない。同様に、

階層木「39」の一部を示したが、rdam:P30327 (contributor agent to aggregate) を最上位エレメントとして、最大5階層を形成している。その中でrdam:P30337 (surveyor person) は、rdam:P30312 (surveyor agent) とrdam:P30338 (contributor person of cartography) という2つのエレメントの下に配置されており、多重階層の事例に該当する。

また、DC タームについては、DCMI Usage Board (2020, January 20) が rdfs:subPropertyOf によって階層関係を定義している。RDA マッピングに出現する範囲では多重階層ではなく、単純な階層構造である。例えば、dct:creator は dct:contributor と http://purl.org/dc/elements/1.1/creator の両者のサブプロパティであるが、後者の基本15要素からなるDCMESプロパティはRDAマッピングにおいては採用されていない。

本研究では、LRM へのマッピングと同様、RDA エレメントの階層木に従ってマッピング行を配置したときに、マッピング先のDCタームが直上位エレメント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集計した(第3表)。

1) 指標 1-1: Code1=0 (最上位レベルのエレメント) のマッピング行が156 (全体の4.7%)、Code1=1 (階層的に整合) が3,114 (93.7%)、Code1=3 (階層的に不整合) が54 (1.6%) であった。Code1=2 (部分的に整合) は発生していない。

第2図において階層木「23」のマッピング先はdct:titleとdct:alternativeの2つであり、加えてdct:alternativeはdct:titleのサブプロパティであり、マッピング先が階層関係にある。それに対して、階層木「39」の図示した範囲においては、マッピング先が階層関係にない事例が出現している。それは「*」を前置したrdam:P30312と、最後のrdam:P30337である。これらはいずれも直上位エレメントのマッピング先dct:contributorとは階層関係にない、dct:publisherをマッピング先としている。図示していない他の不整合となる例には、rdae:P20123 (dramatization of expression) など、そのマッピング先をdct:isVersionOfとす

るのに対して、直上位エレメントrdae:P20152 (adaptation of expression) のマッピング先はdct:hasVersionであり、これら2つのDCタームは逆向きの関連に当たるため、階層関係にはない例などがある。

RDA エレメント階層と整合したマッピング先の割合を表す、指標 1-1 の値 0.984 が得られた。

2) 指標 1-2: 同一 RDA エレメントの複数回出現による整合箇所を考慮した場合、「階層的に整合」が3,127 (94.1%)、「階層的に不整合」が41 (1.2%) となり、指標値 0.988 が得られた。

第2図末尾のrdam:P30337はこれに該当し、「階層的に整合」に移行する例となる。また、rdam:P30366 (surveyor collective agent) などは、図示した箇所ではCode1=1であるが、他の出現箇所ではCode1=3があることを、Code2において示している。

3) 指標 1-3: マッピング先要素の照合を当該 RDA エレメント階層木の最上位エレメントにまで拡張した場合、Code3=1 (階層的に整合) が3,121 (93.9%)、Code3=3 (階層的に不整合) が47 (1.4%) となり、指標 1-3 の値は0.986であった。

図示した範囲ではCode1の値から変更になる事例はないが、マッピング全体において計13箇所、Code1=3に対してCode3=1となる事例が発生している。

以上の通り、照合条件の緩和によって、整合性割合を表す指標値も若干ながら上昇を見せた。ただし、指標 1-2 の値が最も高い結果であった。

C. MARC フィールドへのマッピング

RDA エレメントの階層関係に基づき582の階層木が形成された。このうち、複数階層による木が228、階層を形成しない孤立エレメントが354であった(第2表)。平均階層数は2.0 (SD 1.7)、平均エレメント異なり数14.2 (SD 50.5)、平均マッピング行数57.2 (SD 302.8) であった。最大規模の階層木は、rdae:P20301 [IRI] (related agent of expression) と他の3つの記録方法それぞれからなる計4つのエレメントを最上位とする

Group no.	Hierarchy level	RDA element	Element name	Occurrence count	Hierarchical consistency code: Code1 Code2 Code3	MARC field
Mapping to MARC fields						
115	1	rdaw:P30134 [unstructured description]	title of manifestation	1	0	B245 ** \$a B500 ** \$a
115	2	rdaw:P30128 [unstructured description]	variant title of manifestation	1	3	+B246 ** \$a +B740 ** \$a
115	2	rdaw:P30131 [unstructured description]	abbreviated title	1	2	B500 ** \$a +B210 ** \$a
115	2	rdaw:P30156 [unstructured description]	title proper	1	1	B245 ** \$a
115	3	rdaw:P30203 [unstructured description]	parallel title proper	1	3	+B245 ** \$b
115	4	rdaw:P30204 [unstructured description]	parallel title of series	2	3	+B246 ** \$a +B490 ** \$a
115	2	rdaw:P30157 [unstructured description]	title of series	1	2	B500 ** \$a
115	3	rdaw:P30204 [unstructured description]	parallel title of series	2	1	B490 ** \$a

212	1	rdaw:P10311 [IRI]	related agent of work	1	0	
		A500 ** \$1 A510 ** \$1	A511 ** \$1 B100 ** \$1	B110 ** \$1	B700 ** \$1	B711 ** \$1 B720 (#2)* \$1
212	2	rdaw:P10065 [IRI]	creator agent of work	1	1	
		A500 ** \$1 A510 ** \$1	A511 ** \$1 B100 ** \$1	B110 ** \$1	B700 ** \$1	B711 ** \$1 B720 (#2)* \$1
212	3	rdaw:P10051 [IRI]	designer agent	1	1	
		A500 ** \$1 A510 ** \$1	B100 ** \$1 B110 ** \$1	B700 ** \$1	B710 ** \$1	B720 (#2)* \$1
212	4	rdaw:P10441 [IRI]	designer person	3	2	
		A500 (01)* \$1	B100 (01)* \$1	B700 (01)* \$1		+B720 1* \$1
212	4	rdaw:P10488 [IRI]	designer collective agent	3	1	
		A500 3* \$1 A510 ** \$1	B100 3* \$1 B110 ** \$1	B700 3* \$1	B710 ** \$1	B720 (#2)* \$1
212	5	rdaw:P10535 [IRI]	designer corporate body	6	1	
		A510 ** \$1	B110 ** \$1	B710 ** \$1		B720 (#2)* \$1
212	5	rdaw:P10582 [IRI]	designer family	6	1	
		A500 3* \$1	B100 3* \$1	B700 3* \$1		B720 (#2)* \$1
...						
212	3	rdaw:P10437 [IRI]	creator person of work	2	2	
		A500 (01)* \$1	B100 (01)* \$1	B700 (01)* \$1		+B720 1* \$1
212	4	rdaw:P10436 [IRI]	author person	3	1	
		A500 (01)* \$1	B100 (01)* \$1	B700 (01)* \$1		B720 1* \$1
212	5	rdaw:P10429 [IRI]	screenwriter person	4	1	
		A500 (01)* \$1	B100 (01)* \$1	B700 (01)* \$1		B720 1* \$1
212	4	rdaw:P10441 [IRI]	designer person	3	1	
		A500 (01)* \$1	B100 (01)* \$1	B700 (01)* \$1		B720 1* \$1

第3図 MARC フィールドへのマッピング階層表示

4つの木である。いずれも構成するエレメント異なり数 300, ノード数 (マッピング行数) 2,513, 階層数 9 であった。なお, ここではマッピング先は MARC フィールドの組み合わせを単位としており, 階層木を形成させたときのマッピング行数は合計で 33,316 となった。

第 3 図に, RDA エレメント階層木の一部を示した。階層木「115」は, 最上位エレメント `rdam:P30134 [unstructured]` (title of manifestation) としている。もう一つの階層木「212」は, `rdaw:P10311 [IRI]` (related agent of work) を最上位とし, 総計 297 エレメントが行数 1,204 からなる 7 階層の木を形成している。エレメント `rdaw:P10535 [IRI]` (designer corporate body) などは, 計 6 回ずつ出現していることが出現回数として示されている。

MARC フィールドへのマッピングにおいては, 基本的に複数のフィールドがマッピング先として指定されている。こうした状況の説明図として第 1 図-b) を示した。RDA エレメントが 2 つの階層木を構成し, それぞれのノードとなるエレメントが単一または複数のマッピング先要素をもつ場面である。s1 のマッピング先は {t1, t2, t3} の組み合わせとなり, s2 と s3 も同じマッピング先要素を共有している。他方, s4 のマッピング先は {t3, t4} であり, ここでは t1>t4 の階層関係があるとしている。また, s6 のマッピング先は {t4, t5} であり, RDA エレメントとマッピング先要素の両者の階層関係の観点からは, s6 から t4 へのマッピングは整合的であるが, t5 へのマッピングは整合しない, すなわち「部分的に整合」となる。あるいは, s8 を最上位とする階層木において, 下位エレメント s6 のマッピング先が s8 のそれとは階層的に整合しない。また, ここでは t6>t8 という階層関係があるとしており, そのため s9 からのマッピングは階層的に整合する。

さて, ここで取り上げる MARC フィールド自体に階層関係はない。しかし, それに代えて下記の包含関係を適用できる。

- 1) インディケータの包含関係: 値「1」や「#」(空値) は, 値「*」(任意を表す) に合致し

包含されると見なすことができる。これにより, 「B264*1\$a」は「B264**\$a」に包含される。この処理に加えて, 同一エレメントからのマッピング先のインディケータ値が複数指定されている場合には, 統合化した表示とする (例: 「B264*(1|2|3)\$a」)。

- 2) サブフィールドの包含関係: 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ことができる。例えば, 「A510**\$a, b, c, d, g」に「A510**\$a, c」は包含されると見なせる。マッピング先を個別サブフィールドごとに独立させている場合 (「A111 1*\$a」と「A111 1*\$c」) には, 便宜的に統合化して扱うことにした。
- 3) 単一 RDA エレメントから複数の 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。上記の 1) と 2) を取り込んだ上で, こうした組み合わせにおける包含関係を判定する。直上位エレメントのマッピング先要素が「A510**\$a, c」と「A511**\$a」, 当該エレメントのマッピング先が「A510**\$a, c」のみであるならば, ここには包含関係が認められる。

これらを適用して, 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となる。その場合, 第 1 図-b) に示したような「部分的に整合」という事例も発生する。

第 3 図に示した階層表示では, マッピング先 MARC フィールドは階層木内で同一のものが同じ列となるよう揃えた配置とした。ただし, 同図ではスペースの制約でそれに準じた表示としてある。また, 階層木「212」についてはマッピング先 MARC フィールドの種類数が多く, そのため改行して配置した。別途公開しているファイルでは, こうした改行は挿入しておらず, より見やすいデータであるため, そちらを参照されたい。

- 1) 指標 1-1: Code1=0 (最上位レベルのエレメント) が 582 (全体の 1.7%), Code1=1 (階層的に整合) が 30,142 (90.5%), Code1=2 (部分的に整合) が 2,502 (7.5%), Code1=3 (階層的に不整合) が 90 (0.3%) であった (第 3 表)。

Code1=2 (部分的に整合) の事例には、第3図に示した階層木「115」の rdam:P30131 [unstructured] (abbreviated title) と rdam:P30157 [unstructured] (title of series), 階層木「212」の rdaw:P10441 [IRI] (designer person) と rdaw:P10437 [IRI] (creator person of work) が該当する。rdam:P30131 [unstructured] の場合、マッピング先の「B500**\$a」は直上位エレメントのマッピング先と共通するが、もう1つのマッピング先「B210**\$a」は階層的に整合しない。このような整合しない箇所には「+」を前置して示した。また、Code1=3 (階層的に不整合) の事例には、階層木「115」の rdam:P30128 [unstructured] (variant title of manifestation) など、3箇所が該当している。

「最上位レベルのエレメント」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 となる。

2) 指標 1-2: 当該エレメントの他の出現箇所での整合を示す Code2=1 の場合、Code1=2, 3 を置換できると見なすと、Code1=2 でそれに該当する箇所 1,776, Code1=3 で該当する箇所 20 がそれぞれ見られた。

階層木「115」で rdam:P30204 [unstructured] (parallel title of series) は 2 回出現しており、それぞれ Code1=1 と Code1=3 であるため、Code2 においても一方の値を示している。同様に、階層木「212」において、rdaw:P10441 [IRI] (designer person) も、Code1=1 と Code1=2 として出現しており、Code2 においても一方の値を示している。

このように Code2 の値を組み合わせると、「階層的に整合」となる箇所が合計 31,938 (95.9%) に増加する。また、Code2=2 をもって Code1=3 を置換できる、すなわち「階層的に不整合」から「部分的に整合」に移行する事例が 1 つあり、最終的に「階層的に不整合」のまま残る事例は 69 (0.2%) に減少した。これらゆえ、指標 1-2 の値は、0.976 となった。

3) 指標 1-3: 最上位エレメントのマッピング先 MARC フィールドまで照合範囲を拡張した

場合、Code3=1 (階層的に整合) が 31,303 (94.0%), Code3=2 (部分的に整合) が 1,359 (4.1%), Code3=3 (階層的に不整合) が 72 (0.2%) であった。ここから指標 1-3 の値は 0.957 となり、指標 1-2 の値には届かない結果となった。

第3図において、Code1 の値から変更になる事例はないが、マッピング全体において 1,145 箇所 Code1=2 に対して Code3=1 となる事例が発生し、16 箇所 Code1=3 に対して Code3=1 となる事例が発生している。

RDA エレメントベースの指標値についてまとめると、LRM, DC, MARC21 へのマッピングにおいては、いずれも RDA エレメント階層とマッピング先要素の階層とは 90% 以上の整合性を保っているとはいえ、逸脱する事例が一定数見られる結果となった。照合条件を緩めると整合となる事例が増えるが、依然として逸脱事例は残される。

IV. RDA エレメント階層木ベースの指標

A. エレメント階層木ベースの指標の概要

2 つめの指標群は、RDA エレメント階層木ベースの指標とし、階層木単位の RDA グラフ分割と、引用分析における書誌結合を流用したマッピング先の共有に基づく書誌結合による RDA グラフ分割とが、どの程度類似したものとなるかを示す指標とする。2 つのグラフ分割結果が異なるほど、マッピングは複雑となる。

RDA エレメント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。階層木内での同一エレメントの複数回出現は単一化されるが、同一エレメントが複数の階層木にまたがって出現するものはそのまま多重出現となる。他方、マッピング先の共有に基づく書誌結合とは、RDA エレメントとマッピング先からなる 2 部グラフから、RDA エレメントへのプロジェクションをとり RDA グラフ分割を行うことに等しい。これら階層木によるグラフ分割結果であるサブグラフ (クラスタ) と書誌結合によるサブグラフとが一致する状況とは、大まかにいえ

ば RDA エLEMENT の階層構造とそれに整合したマッピング先要素が指定されている、すなわち同一（または階層関係にある）要素が指定された状況である。前章では RDA エLEMENT ごとのマッピング先の整合性を見てきたが、本章ではELEMENT 階層木を単位とするELEMENT の集合によって整合性を測るという相違がある。グラフ分割結果の類似度は、その上限（例えば「1」）から減じることによって、複雑さの指標値となる。

1. ELEMENT ペアベース精度・再現率・F 値

RDA ELEMENT 階層木によるグラフ分割を正解と見なした場合に、書誌結合による RDA グラフ分割結果について、ELEMENT のペアベースで精度・再現率・F 値 (F1 score) を求める方法を採用した (Menestrina et al., 2010)。同一ELEMENT の複数サブグラフ (クラスタ) における出現すなわち多重出現に対処できる指標を選択した。

第 1 図-a) の例を用いれば、ELEMENT 階層木から得られる正解クラスタは 2 つとなり、それぞれ {s1, s2, s3, s4, s5, s6, s7} と {s6, s8, s9} の ELEMENT 群から構成される (s6 が多重出現)。先頭のクラスタから ELEMENT ペア s1-s2, s1-s3 など 21 ペア、2 つめのクラスタからペア s6-s8, s6-s9, s8-s9 の 3 ペアが導かれ、総計 24 ペアが正解ペアとなる。一方、書誌結合から得られる予測クラスタは、{s1, s2, s3}, {s4}, {s5}, {s6}, {s7, s8}, {s9} と細分化されたものとなり、ここからは s1-s2, s1-s3, s2-s3, s7-s8 の計 4 ペアのみ導かれる。{s4}, {s5} などの孤立した ELEMENT によるクラスタからは、ペアは形成されないからである。

そして、正解と予測で一致するペアすなわち「正しく予測されたペア」は、s1-s2, s1-s3, s2-s3 の 3 ペアとなり、精度 = (正しく予測されたペア数) / (予測ペア数) = 3/4 = 0.75, 再現率 = (正しく予測されたペア数) / (正解ペア数) = 3/24 = 0.125, そして F 値 = 0.214 となる。多重出現の ELEMENT s6 は、形成するペアをそれぞれカウントすることで正しく処理される。この事例では正解クラスタが大きい分、そのペア数も多くなり、再現率は低い

値となる。他方、少数とはいえ、正解ペアを的確に予測していたため精度は高い結果となる。

次に、マッピング先要素の階層関係を組み入れて書誌結合をとった場合、予測クラスタは {s1, s2, s3, s4, s6}, {s5}, {s7, s8, s9} となり、先頭クラスタからは s1-s2, s1-s3 など 10 ペア、最後のクラスタからは s7-s8, s7-s9, s8-s9 の 3 ペアが導かれる。正解と予測で一致するペアは、先頭の予測クラスタの 10 ペアすべてと、最後の予測クラスタから s8-s9 のみ (s7-s8 と s7-s9 は正解には出現しない) となり、計 11 ペアとなる。よって、精度 = 11/13 = 0.846, 再現率 = 11/24 = 0.458, F 値 = 0.594 であり、いずれも上昇する結果となる。

RDA ELEMENT からのマッピング先要素が複数となる場合 (第 1 図-b)) についても、マッピング先の個別要素に対する書誌結合によって、グラフ分割を行う。同図において書誌結合による予測クラスタは {s1, s2, s3, s4, s5, s6}, {s7, s8}, {s9} の 3 つであり、マッピング先要素 t3 と t4 によって最初のクラスタが形成されている。正解クラスタは第 1 図-a) と同じであるため、先頭の予測クラスタから導かれる 15 ペアはすべて正解クラスタからのペアに含まれる。しかし、2 つめの予測クラスタによるペア s7-s8 は、正解クラスタからのペアには含まれていない。よって、精度 = 15/(15+1) = 0.938, 再現率 = 15/24 = 0.625, F 値 = 0.750 となる。また、マッピング先の階層関係 t1>t4, t6>t8 を組み入れた書誌結合の場合、予測クラスタは {s1, s2, s3, s4, s5, s6}, {s7, s8, s9} に更新される。先頭の予測クラスタには変更はなく、2 つめの予測クラスタからは s7-s8, s7-s9, s8-s9 の 3 ペアが導かれ、そのうち s8-s9 が正解クラスタからのペアに含まれる。よって、正しく予測されたペア数 16 となり、精度 = 16/(15+3) = 0.889, 再現率 = 16/24 = 0.667, F 値 = 0.762 と求められる。この事例では、マッピング先の階層関係を組み入れた結果、精度は低下し、再現率と F 値は上昇している。

2. Omega Index と双方向 Mean F1 score

正解をもたないグラフ分割結果の類似性を求め

る指標である、Omega Index と双方向 Mean F1 score (Two-way Mean F1 score, 双方向平均 F 値)。以後、MF1 と略記) を加えて採用した (Murray et al, 2012; Lutov et al, 2019)。Omega Index は、同一要素の多重出現すなわちオーバーラップクラスタに適合した評価指標である。それに対して、MF1 は厳密には多重出現に対応したものではないが、同一事象を測る異なる指標として採用した。

Omega Index は、エレメントペアに着目し、各ペアがそれぞれの分割において共有するクラスタ数に基づいて、2つの分割結果の一致度を評価する指標である。すべてのペアについて、まず2つの分割結果において当該ペアが出現するクラスタ数 (すなわち、いくつのクラスタにおいてペアを形成するか) を数え、次に両方の分割結果でそのクラスタ数が一致したペアを数え、最後に一致ペア数の全ペア数に対する割合を求める。値は0から1の範囲をとり、1に近いほど2つの分割結果が類似していることを表す。例えば第1図-a) に示した事例の場合、ノード s1~s9 のすべてのペアは36となり、個別ペア s1-s2 は両方の分割結果においていずれも1つずつのクラスタに出現し、ペア s1-s4 は階層木からのクラスタには出現するが、書誌結合によるクラスタには出現しない (共有クラスタ数0)。このようにすべてのペアについて、共有クラスタ数が一致するペアを数える。この事例では12ペアが一致し、Omega Index = $12/36 \approx 0.333$ となる。また、マッピング先の階層関係を組み入れて書誌結合を取った場合には、Omega Index ≈ 0.611 となる。なお、Omega Index には、「偶然の一致」を考慮した補正を加えたものとして行うことができるが、ここでは検討の見通しをよくするため単純な指標を採用する。

一方の MF1 は、まず分割結果1の個々のクラスタに対して、構成する要素の共通性において分割結果2の中で最もよく対応する (すなわち F 値が最大の) クラスタを求め、次に全クラスタのこうした F 値の平均を取ることで、分割全体の類似度を評価する (単方向 MF1)。次に、方向を逆転させ、分割結果2の各クラスタに対して、分割結果1の中で最も F 値が高いクラスタを探

し、全クラスタについてこの F 値の平均を取る。最後に、こうして得た2つの平均 F 値のさらに平均を求める (双方向 MF1)。この方法により、局所的なクラスタ単位で、対応する精度と再現率のバランスを考慮した類似度評価が可能となる。MF1 の値も0から1の範囲をとり、1に近いほど分割結果が類似していることを表す。

例えば第1図-a) の事例において分割1 (階層木による分割) によるクラスタ {s1, s2, s3, s4, s5, s6, s7} に対して、分割2のクラスタ群のうち、最も F 値が高いのはクラスタ {s1, s2, s3} である。そこでは精度 = $3/3 = 1.0$ 、再現率 = $3/7 \approx 0.429$ 、F 値 ≈ 0.600 となる。分割1の2つめのクラスタ {s6, s8, s9} については、分割2のクラスタ {s6} と {s9} がいずれも精度 = $1/1 = 1.0$ 、再現率 = $1/3 \approx 0.333$ 、F 値 ≈ 0.500 である。これらから、単方向 MF1 $\approx (0.600 + 0.500)/2 = 0.55$ である。次に、逆方向で処理を行い、分割2の個々のクラスタに対して、最大の F 値となる分割1のクラスタを探し、それらの F 値の平均を求める。その結果、単方向 MF1 = $(0.6 + 0.25 + 0.25 + 0.5 + 0.4 + 0.5)/6 \approx 0.417$ が得られる。最終的には、これら2つの単方向 MF1 の平均である、双方向 MF1 ≈ 0.483 が求める指標値となる。なお、分割2にマッピング先の階層関係を組み入れて書誌結合を取った場合、すなわち分割2を {s1, s2, s3, s4, s6}, {s5}, {s7, s8, s9} としたときには、MF1 ≈ 0.683 となる。

B. LRM プロパティと DC タームへのマッピング

LRM プロパティへのマッピングでは、a) RDA エレメント階層木による分割がクラスタ数149、そのうち最大エレメント数をもつクラスタのエレメント数2,739、クラスタごとの平均エレメント数20.3 (SD 224.3) であり、エレメント数の分布のジニ係数は0.94という極めて偏りの大きな分割結果となっている (第4表)。同様に、b) 書誌結合によるグラフ分割では、クラスタ数103、クラスタごとの平均エレメント数29.4 (SD 74.0)、分布のジニ係数0.80となった。エレメント階層木によるグラフ分割を正解として、書誌結合による分割結果を評価したところ、ペアベース精度0.979、再現

第4表 RDA エLEMENTのグラフ分割

	クラスタ数	最大 ELEMENT数	平均ELEMENT数 (SD)	ジニ 係数
LRM プロパティへのマッピング				
a) RDA ELEMENT階層木による分割	149	2,739	20.3 (224.3)	0.94
b) 書誌結合による分割	103	509	29.4 (74.0)	0.80
c) 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	2,725	84.0 (453.2)	0.95
DC タームへのマッピング				
a) RDA ELEMENT階層木による分割	156	293	7.5 (31.0)	0.80
b) 書誌結合による分割	32	360	36.0 (74.2)	0.75
c) 書誌結合による分割かつ DC タームの階層性組み入れ	22	360	52.4 (92.8)	0.73
d) RDA 階層木かつ階層レベル調整による分割	96	293	12.2 (41.9)	0.83
MARC フィールドへのマッピング				
a) RDA ELEMENT階層木による分割	582	300	14.2 (50.5)	0.86
b) 書誌結合による分割	165	3,723	50.0 (366.3)	0.96
c) 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	3,743	55.1 (385.5)	0.96
d) RDA 階層木かつ階層レベル調整による分割	530	300	15.6 (52.8)	0.86
e) 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	494	7.2 (25.1)	0.77

第5表 RDA ELEMENT階層木ベースの指標値（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	双方向 Mean F1 score
LRM プロパティへのマッピング					
a) RDA ELEMENT階層木による分割	.979	.084	.155	.029	.238
- b) 書誌結合による分割					
a) - c) 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	.990	.982	.986	.923	.412
DC タームへのマッピング					
a) RDA ELEMENT階層木による分割	.701	.968	.813	.783	.404
- b) 書誌結合による分割					
a) - c) 書誌結合による分割かつ DC ター ムの階層性組み入れ	.646	.988	.781	.744	.425
d) RDA 階層木かつ階層レベル調整による 分割 - b)	.752	.908	.822	.791	.441
MARC フィールドへのマッピング					
a) RDA ELEMENT階層木による分割	.071	.999	.132	.093	.479
- b) 書誌結合による分割					
a) - c) 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	.070	.999	.132	.091	.500
d) RDA 階層木かつ階層レベル調整による 分割 - b)	.071	.998	.133	.093	.472
a) - e) 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	.471	.229	.308	.297	.432
d) - e)	.475	.230	.309	.299	.430

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 (第 5 表)。

次に, c) 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, すなわちマッピング先 LRM プロパティに階層関係があるときには当該エレメントのマッピング先プロパティに加えてその上位プロパティをもマッピング先とした場合, クラスタ数は 36 に大きく減少し, その分, クラスタの平均エレメント数は 84.0 (SD 453.2) に増加した。a) 階層木による分割に対する, この c) LRM プロパティの階層性を組み入れた上での書誌結合による分割は, 精度 0.990, 再現率 0.982, F 値 0.986 をもたらした。再現率の改善によって, 高い F 値が得られている。同様に, a) と b) との分割類似度 Omega Index と MF1 の値は 0.029 と 0.238 であったものが, a) と c) の場合には 0.923 と 0.412 という, 大きな上昇を見せている。

一方, DC タームへのマッピングについては, a) RDA エレメント階層木による分割はクラスタ数 156 であり, 最大エレメント数 293, 平均エレメント数 7.5 (SD 31.0) という, 比較的小規模なクラスタ群を形成した (第 4 表)。それに対して, b) 書誌結合による分割ではクラスタ数 32, 平均エレメント数 36 (SD 74.2) という, 大きなクラスタ群となった。a) に対する b) の分割性能は, 精度 0.701, 再現率 0.968, F 値 0.813 という高い値であった (第 5 表)。

c) 書誌結合を取る際に DC タームの階層性を組み入れると, クラスタ数は 22 に減少した (平均エレメント数 52.4, SD 92.8)。そのため, a) に対する c) の分割性能は, 精度 0.646, 再現率 0.988, F 値 0.781 となり, 僅かとはいえ精度および F 値が低下している。

また, d) RDA 階層木による分割を行う際に, 個別の階層木の最上位エレメントに対して, さらに 1 段上位のエレメントのレベルで合致した階層木を統合するという処理を加えた分割を準備した。DC タームへのマッピングに出現する RDA エレメントは, 定義されたエレメント全体の階層構造において中位以下に位置するものに限定され

ている点を考慮した処置である。例えば, 最上位エレメント rdaw:P10002 (identifier for work), rdaw:P10088 (title of work), rdaw:P10328 (access point for work) は, その直上位がマッピングに未出現の rdaw:P10329 (appellation of work) であり, これら 3 エレメントを最上位とする階層木を単一階層木に統合した。このような処置の結果, クラスタ数は 96 に減少した (平均エレメント数 12.2, SD 41.9)。このグラフ分割結果を正解として, b) 書誌結合による分割の性能を測定したところ, 精度 0.752, 再現率 0.908, F 値 0.822 となり, 先の a) を用いたときに比べて, 精度と F 値が多少とも上昇し, 再現率は低下する結果となった。F 値を全体的な評価値とすれば, DC タームへのマッピングについては, d) RDA エレメントの階層レベル調整による階層木統合は, b) 書誌結合による分割結果に近づくことになる。Omega Index と MF1 の値も, 同じ傾向を示している。

C. MARC フィールドへのマッピング

a) RDA エレメント階層木による分割は, クラスタ数 582, クラスタの最大エレメント数 300, 平均エレメント数 14.2 (SD 50.5) であった (第 4 表)。それに対して, b) 個別 MARC フィールド単位での書誌結合による分割は, クラスタ数 165 で, 最大エレメント数 3,723, 平均エレメント数 50.0 (SD 366.3) であった。マッピング先として極めて多数回出現する MARC フィールド (例えば「B500**\$a」など) が巨大なクラスタを導く一方, 極小のクラスタも多数存在するという, 振りが大きい分割である (ジニ係数 0.96)。a) に対する b) の分割性能は, 精度 0.071, 再現率 0.999, F 値 0.132 となり, 精度が極端に低く, 再現率が極めて高い値となった (第 5 表)。

次に, c) マッピング先 MARC フィールドの階層的包含を組み入れた上で書誌結合による分割を行った場合, クラスタ数 150, 平均エレメント数 55.1 (SD 385.5) となり, 大きな変化は見られなかった。そのため, a) に対する分割性能につ

いても、殆ど変化が見られない。

d) RDA 階層木について、最上位エレメントのさらに一段上位エレメントが複数の階層木において合致したとき、あるいは一方の階層木の最上位エレメントの一段上位エレメントが他の階層木の最上位エレメントの直下位に位置づけられるとき、それら階層木を統合するという処置を試みた。最上位エレメント rdaa:P50121 [IRI] (date of birth), rdaa:P50120 [IRI] (date of death) などの直上位エレメント rdaa:P50347 [IRI] (related timespan of person) は、他の階層木の最上位エレメント rdaa:P50339 [IRI] (related timespan of agent) の下位エレメントであり、これらの階層木を統合するといった処置である。しかし、統合によってもクラスタ数は 530 (平均エレメント数 15.6, SD 52.8) となり、あまり大きな変更とはならなかった。これらゆえ、d) に対する b) の分割性能も、元の a) による分割を用いた場合と、殆ど変化が見られない。

同様に、Omega Index と MF1 の値についても、a) と b) の分割類似度、a) と c) の類似度、d) と b) の類似度は、いずれも Omega Index が 0.1 以下、MF1 が 0.47 から 0.50 という値となった。

加えて、e) 書誌結合を取るマッピング先要素を、個別 MARC フィールド単位ではなく、MARC フィールドの組み合わせを単位として実行した。第 1 図-b) の事例を用いれば、この方式では書誌結合から得られるクラスタは {s1, s2, s3}, {s4, s5}, {s6}, {s7, s8}, {s9} となる。こうした分割の結果、クラスタ数 1,145、平均エレメント数 7.2 (SD 25.1) となり、b) のフィールド単位の書誌結合による分割とは大きく異なるものとなった。これを用いて、a) の階層木による分割との類似度を計測すると、精度 0.471、再現率 0.229、F 値 0.308 となり、値は向上した。さらに、d) 階層木に階層レベル調整による統合を加えた分割とこの e) による分割との類似度は、精度 0.475、再現率 0.230、F 値 0.309 であり、僅かながらも上昇した。a) と e) の分割類似度、d) と e) の類似度は、それぞれ Omega Index が 0.297 と 0.299、MF1 が 0.432 と 0.430 となり、

Omega Index については a) と b) との類似度に比べて向上している。

V. おわりに

本研究は、RDA から IFLA LRM, DC メタデータ語彙, MARC21 へのマッピングという、大規模かつ複雑なマッピングに対して、a) 基本的な集計によってその概要を提示し、また b) それらマッピングを複雑にしている主たる要因とその状況を示し、さらに c) マッピングの整合性を適切に表す指標群を導入し、実際に指標値を求めることを行った。整合性の指標値は、その上限から減じることによって複雑性の指標値とすることもできる。

RDA エレメントは階層性をもち、しかも上位エレメントを複数もつという多重階層として定義されている。また、マッピング先の要素もそのスキーマに依存して階層性を備えており、さらにはマッピング先要素が複数である場合には部分集合としての包含関係が発生する。こうしたマッピング元の RDA エレメント階層構造と、マッピング先の階層関係・包含関係が整合的となるかを、マッピングの整合性の要因として捉え、それに関する指標を採用した。

RDA エレメントベースの指標には、マッピング元とマッピング先の両者における階層関係がどの程度整合するのかを、該当する事例数の割合をもって示した。3つのスキーマへのいずれのマッピングにおいても整合する事例は 90% を超えていることを確認した。また、その整合性判定の条件を緩和することによって、不整合となる事例数を減少させることができることを併せて示した。ただし、3つのいずれのマッピングにおいても、かついずれの条件緩和においても整合しない事例が一定数残された。

また、RDA エレメント階層木ベースの指標として、階層木による RDA グラフ分割とマッピング先の書誌結合による RDA グラフ分割とが、どの程度類似したものとなるかを複数の指標によって示した。これら 2つの分割結果が異なるほど、マッピングが構成する構図は複雑となる。さら

に、マッピング先の階層関係を組み入れた上での書誌結合、RDA エlement階層木における最上位Elementのレベル調整による階層木統合などを加えた場合についても同様に指標値を求めた。

これら2群の指標によって、3つのマッピングそれぞれの整合性と、それらマッピング間の指標値から見た相対的な位置関係を確認できた。

なお、マッピング元と先の両者の階層性から逸脱する事例も相当数残される結果となり、それらは個別Elementまたは階層木とマッピング先データ要素に固有の事情を確認することによってのみ、不整合となる理由を理解することができる。

最後に、RDA 運営委員会には、マッピングの公開に加えて、その策定方針を説明した文書などの公開を期待したい。マッピングは策定の意図によって、同じマッピング先スキーマに対するものであっても、大きく異なる内容となることが考えられるからである。特に、マッピング先が多数設定されている MARC21 へのマッピングなどでは、その採用範囲や基準などに依存して、異なるマッピングが導かれる可能性がある。

付録：第2図、第3図にその一部を示したマッピング階層表示データの全体をファイルとして、下記 URL において公開している。
<https://github.com/ShoichiTaniguchi/RDA-mappings>

注

- 1) 当該バージョン (RDA-Vocabularies. Release v5.1.0, Jun 29, 2024) のマッピングデータには、次の URL でアクセス可能。
<https://github.com/RDAREgistry/RDA-Vocabularies/releases/tag/v5.1.0>

引用文献

- Babalou, S., Kargar, M. J., & Davarpanah, S. H. (2016). Large-scale ontology matching: A review of the literature. In *2016 Second International Conference on Web Research (ICWR)*, Tehran, Iran (pp. 158–165). IEEE. <https://doi.org/10.1109/ICWR.2016.7498461>
- DCMI Usage Board (2020, January 20). *DCMI Metadata Terms*, 2020-01-20. Dublin Core. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Euzenat, J., & Shvaiko, P. (2013). *Ontology Matching* (2nd ed.). Springer.
- Gaitanou, P., Andreou, I., Sicilia, M. S., & Garoufallou, E. (2024). Linked data for libraries: Creating a global knowledge space, a systematic literature review. *Journal of Information Science*, 50(1), 204–244. <https://doi.org/10.1177/01655515221084645>
- Hashhofer, B. (2008). *A comparative study of mapping solutions for enabling metadata interoperability* (TR-20080202). University of Vienna. <https://eprints.cs.univie.ac.at/295/1/covered.pdf>
- IFLA (2020). *IFLA element sets. LRMer*. iflstandards.info. <https://www.iflstandards.info/lrm/lrmer.html>
- Library of Congress (1999–). *MARC standards*. Library of Congress. <https://www.loc.gov/marc/>
- Lutov, A., Khayati, M., & Cudré-Mauroux, P. (2019). Accuracy evaluation of overlapping and multi-resolution clustering algorithms on large datasets. In *Proceedings of IEEE International Conference on Big Data and Smart Computing, BigComp 2019*. <https://doi.org/10.1109/BIGCOMP.2019.8679398>
- Menestrina, D., Whang, S. E., & Garcia-Molina, H. (2010). Evaluating entity resolution results. *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, 3(1-2), 208–219. <https://doi.org/10.14778/1920841.1920871>
- Murray, G., Carenini, G., & Ng, R. (2012). Using the Omega index for evaluating abstractive community detection. In *Proceedings of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization* (pp. 10–18). Association for Computational Linguistics. <https://aclanthology.org/W12-2602.pdf>
- Osman, I., Yahia, S. Y., & Diallo, G. (2021). Ontology integration: Approaches and challenging issues. *Information Fusion*, 71, 38–63. <https://doi.org/10.1016/j.inffus.2021.01.007>
- Otero-Cerdeira, L., Rodríguez-Martínez, F. J., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, 42(2), 949–971. <https://doi.org/10.1016/j.eswa.2014.08.032>
- RDA Steering Committee (2017a). *RDA to MARC Bibliographic mapping: RDA elements mapped to MARC 21 Format for Bibliographic Data*. RDA Toolkit. <https://original.rdatoolkit.org/document.php?id=jscmap1>

- RDA Steering Committee (2017b). *RDA to MARC Authority mapping: RDA elements mapped to MARC 21 Format for Authority Data*. RDA Toolkit. <https://original.rdatoolkit.org/document.php?id=jscmap3>
- RDA Steering Committee (2020-). *Maps for use with RDA vocabularies*. RDA Registry. <https://www.rdaregistry.info/Maps/>
- RIMMF Project (2022-). RIMMF3 Home. <https://rimmf.com/w/doku.php?id=rimmf3:start>
- RIMMF Project (2024-). RIMMF6. <https://rimmf.com/w/doku.php?id=rimmf6:start>
- Scharffe, F., Zamazal, O., & Fensel, D. (2014). Ontology alignment design patterns. *Knowledge and Information Systems*, 40(1), 1-28. <https://doi.org/10.1007/s10115-013-0633-y>
- Taniguchi, S. (2025). Addressing and reducing complexity in metadata mappings: RDA to DC and MARC 21. *Cataloging & Classification Quarterly*, 63(4), 240-266. <https://doi.org/10.1080/01639374.2025.2511804>
- Thiéblin, É., Haemmerlé, O., Hernandez, N., & Trojahn, C. (2019). Survey on complex ontology matching. *Semantic Web*, 11(4), 689-727. <https://doi.org/10.3233/SW-190366>
- University of Washington Libraries (2021-). *MARC2RDA: Mapping and transformation between MARC21 and RDA-RDF*. GitHub. <https://github.com/uwlib-cams/MARC2RDA>
- Woodley, M. S. (2016). Metadata matters: Connecting people and information. In M. Baca (Ed.), *Introduction to Metadata* (3rd ed.). Getty Research Institute. <https://www.getty.edu/publications/intrometadata/metadata-matters/>
- Zeng, M. L., & Qin, J. (2022). 4.5 Crosswalks, and 8. Achieving interoperability. In *Metadata* (3rd ed., pp. 207-211, 355-387). ALA Neal-Schuman.

要 旨

【目的】 RDA から IFLA LRM, DC メタデータ語彙, MARC21 への大規模かつ複雑なマッピングに対して、それを複雑にしている主たる要因とその状況を示し、さらにマッピングの整合性を適切に表す指標群を導入し計算することを目的とする。これらによって、マッピングの全体像理解に資することを意図している。

【方法】 RDA 語彙 v5.1.0 において示された、3つの対象スキーマそれぞれへのマッピングについて基本的な集計を行った。次に、マッピング元の RDA エレメント階層構造（ただし、多重階層）とマッピング先要素の階層関係・包含関係が整合的となるかを、マッピングの整合性の要因として着目した。これらの状況を適切に表し評価するために、a) RDA エレメントベースの指標群と、b) RDA エレメント階層木ベースの指標群の2種類を導入した。

【結果】 基本的集計によって、3つのマッピングの規模を表す RDA エレメント数、マッピングペア数はかなり異なる点を示した。次に、a) 3つのマッピングすべてにおいて、RDA エレメントとマッピング先要素の両者における階層関係は90%以上が整合していた。また、整合性判定の条件を緩和することによって、不整合となる事例を減少させることができたが、依然として整合しない事例が一定数残された。b) 階層木による RDA グラフ分割とマッピング先の書誌結合による RDA グラフ分割とが類似する程度を複数の指標（エレメントペアベースの精度・再現率・F 値, Omega Index など）によって示した。さらに、マッピング先の階層関係を組み入れた上での書誌結合、RDA エレメント階層木における最上位エレメントのレベル調整による階層木統合などを加えた場合について同様に指標値を求め、その効果を確認できた。これらの指標群によって、3つのマッピングそれぞれの整合性とその相対的な位置関係を確認できた。