

RDA 規定の RDF データ化

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

橋詰 秋子 (実践女子大学短期大学部)
hashizume-akiko@jissen.ac.jp

RDA のインストラクションおよびガイダンスの規定を対象にして、主に活用の観点から適切な RDF データ表現とすることを検討した。RDA 語彙の RDF 定義済みデータからの接続を前提にした上で、①URI 付与単位の設定と付与単位間の関係表現、②規定内の指示内容の表現、③参照指示や例示の扱い、④適用細則である *policy statements* の表現と RDA 規定との接続などについて検討した。RDA Toolkit サイト上から規定データをスクレイピングによって RDF データに機械的変換し、その後に人手による確認と修正を加え、提案方式の妥当性を検証した。

1. はじめに

筆者らはこれまでに、メタデータ記述規則の機械可読化、特に RDF データ化を意図して、「日本目録規則 2018 年版 (NCR2018)」の記述規則を RDF データ表現とすることを検討した。検討に当たっては、記述規則とそれを適用して作成されたメタデータとを接続する、規則内外の参照関係を幅広く迎えるようにするといった活用の観点を重視した。既発表の研究では、NCR2018 語彙の RDF 定義からの接続を前提にした上で、①個別規定の設定単位および関係表現、②個別規定から当該規定に従って作成されたメタデータへの参照、③逆にメタデータから実際に適用された個別規定への参照などについて、それぞれ必要な検討項目を示し、可能な選択肢を提示した¹⁾。併せて、NCR2018 の一部の規定群を対象に、実際に RDF データ化を試行した。

本発表では、上記の成果を踏まえつつ、RDA (Resource Description and Access) を取り上げ、その規定群を適切な RDF 表現とする方式を検討し提案する。検討は NCR2018 規定群の RDF 化と基本的に同じ視点に立ちつつも、RDA 特有の事項を考慮する方針とする。併せて、実際に規定群を RDF データに変換し、提案方式の妥当性を検証する。

RDA は、AACR2 の後継として 2010 年に公開を開始したが、改訂が継続して実施されており、現在では RDA Toolkit というウェブサイト (有料) 上でのみ最新版が提供されている。加えて、RDA の再構築を目指す 3R プロジェクトが 2020 年末に完了したことにより、その構成等を大きく変化させている。概念モデル IFLA LRM への対応、新規の概念・事項の導入などがなされた結果、現行の RDA Toolkit は「ガイダンス」、「実体」、「リソース」から構成されている。「ガイダンス」は、RDA の目的・

背景情報、そして複数のエレメントに適用される規定群などをまとめており、いわば総則に該当する。13 個の「実体」と各実体に対する「エレメント」(3,000 以上) の下には、それぞれをどのように扱いメタデータとして記録するかを詳細に定めたインストラクションが示されている。また、「リソース」には、記録に用いる値語彙や適用細則である *Policy statements* などが含まれている。RDA そして RDA Toolkit は、引き続き改訂がなされている。並行して、実体やエレメント、値語彙等の RDF 定義データは、RDA Registry から提供されている。

本研究では、様々な規定群や語彙等からなる RDA を総体として RDF データ化するため、その全体をレイア構造で捉える。レイア 1 は既に RDA Registry 等により URI が付与され定義データが存在する実体、エレメント、値語彙などからなる層、レイア 2 はガイダンスとインストラクションの RDF データを位置づける層、レイア 3 は個別機関が策定する適用細則とアプリケーション・プロファイルなどを扱う層とする。レイア 1 に位置づけられるエレメント等の定義データが、レイア 2 となる個別規定の RDF クラス (のインスタンス) の土台となり、参照されると捉える。

2. 個別規定の RDF 化と規定間の関係指示

a) URI 付与単位と関係指示

実体およびエレメントのインストラクションやガイダンスを構成する個別規定を、RDF クラスのインスタンスとして表現する。現行の RDA では以前にあった条項番号が廃止されており、それに依拠することはできない。クラスを設定する単位 (URI 付与単位) には複数の選択肢があるが、本研究では 1)各インストラクション/ガイダンス内で「見出し」が付いている部分 (「セクション」と呼ぶ) の単位、2)単一

セクションが意味的に独立した規定として扱える複数の部分に分割できる場合にはその最小単位を採用する(第1図参照;詳細は後述のc節)。また、単純な構造のRDFデータとする方針とし、空白ノードによる構造化表現は例示についてのみ適用し、それ以外では採用しない。

セクション間の階層は多段階の構成を取り得るものであり、本研究では独自のプロパティ「:hasSection」と「:isSectionOf」を用いて関係を表現する。RDFデータでは直上位のセクション、直下位のセクションのみ指示し表現する。他方、セクション内が複数の独立した規定に分割されるときには、全体部分関係としてプロパティ「:hasPart」と「:isPartOf」を用いて関係を表現する。この関係性はセクションに該当する全体とその部分という2階層のみの構成とし、多段階の構成とはしない。

b) URIの構成法

現行のRDAは、条項番号ではなく、citation numberと呼ばれるランダムに付与された番号(例:「#21.93.72.57」)が表示される。ただし、付与単位の方針などが不明であるため、今回採用していない。

本研究では独自にURIを付与する。ここには複数の付与方式が考えられる。本研究では、個々の実体とエレメントのインストラクション/ガイダンスごとにURIを付与し、その下の個々のセクションや最小単位の規定はフラグメントURI(ハッシュURI)とする。例えば、エレメント「title proper」のインストラクションURIは、RDAおよびRDA Registryによる当該エレメントのURI「rdam:P30156」(http://rdaregistry.info/Elements/m/P30156)に「i」を付けて「rdam:P30156i」とする。そして、インストラクションを構成するセクションには下記のようなURIを機械的に付与する。

rdam:P30156i#00 ("title proper")
rdam:P30156i#01 ("Definition and Scope")
rdam:P30156i#02 ("Prerecording")
rdam:P30156i#03 ("Recording")
rdam:P30156i#03.01
("Recording an unstructured description")
rdam:P30156i#03.02
("Two or more titles of manifestation")
rdam:P30156i#03.03 ("Titles proper of
manifestations of parts and iterations")

さらに、セクション内最小単位に分割される場合、上記のrdam:P30156i#03.01の下では、rdam:P30156i#03.01.01, rdam:P30156i#03.01.02, rdam:P30156i#03.01.03などのURI、同様にrdam:P30156i#03.02の下では、

rdam:P30156i#03.02.01, rdam:P30156i#03.02.02などのURIを付与する。

なお、URIは「.」(ドット)により階層的な構成に似せているが、必ずしも論理的な階層構造を正しく反映した構成とはしていない。論理的な階層性は前述した関係表現のプロパティ群によって別途表現しているからである。

c) 個別規定の指示内容の表現

RDAの個別規定は、そのメタデータ作成・記録行為の指示を、パターン1)指示する内容をそのまま文として記載する形式、2)見出し「OPTION」に続けて指示内容を記載する形式、3)見出し「CONDITION」の下に条件を記載し、その下に単一または複数のOPTIONを配置する形式で表現している。従来の目録規則でいえば、1)は本則、2)は任意規定または別法、3)は条件付き任意規定または別法に相当する。

本研究では、パターン1)はプロパティ「:instruction」とそのリテラル値で表現する。独立した段落ごとに先のプロパティを適用するため、単一規定(規定リソースURI)に対して複数回の同一プロパティ適用があり得る。なお、「:instruction」と、後掲のプロパティ「:option」や「:condition」とを同一規定内では原則として共起させないが、一部の例外を設けた(後述のパターン4)。

パターン2)は、プロパティ「:option」を適用して指示内容をリテラル値で表す。個別のOPTIONごとに独立した規定とするが、単一OPTION内が複数の段落から構成されているときには、同一規定に対してプロパティ「:option」を繰り返す。また、「:instruction」または「:option」をもつ規定が、同時に「:hasPart」(部分)をもつてはならないとした。必要な場合には、「:instruction」・「:option」自体の箇所も独立させて「部分」とする方針とした。他方、「:instruction」や「:option」をもつ規定が、同時に「:hasSection」(セクション)をもつことは許容する。

パターン3)は、CONDITIONの下にOPTIONごとに独立した規定としてURIを付与し、プロパティ「:condition」と「:option」を組み合わせて表現する。同一CONDITIONの下に複数のOPTIONが並置されている場合には、OPTIONの数だけそれぞれを独立させた規定とする。

加えて、パターン4)指示規定文が"The following option is recommended."のみ、それにOPTIONの規定文から構成されている場合、両者を独立した規定とはせず、組み合わせて単

一規定とした。これは例外的な扱いである。

また、「Prerecording」といった見出しのみ記載があり、その下に具体的な指示内容がない場合にも、当該セクションに独立した URI を付与して RDF データ化した。

d) 他のインストラクションなどへの参照指示

個別規定内で出現した他のインストラクションなどへの参照指示は、下記のプロパティを用いて表現した。

- ・「:referredInstrect」他のエレメントのインストラクションへの参照指示を示す。
- ・「:referredEntyInstrect」他の実体のインストラクションへの参照指示を示す。
- ・「:referredGuidance」ガイダンスへの参照指示を示す。

インストラクション／ガイダンス自身内の「目次」に相当する参照指示は、RDF で表現しない。それらは「:hasSection」によって別途表現されているからである。他方、「目次」以外のインストラクション／ガイダンス内の自己参照はそのまま参照指示として表す。

- ・「:referredResource」RDA 外部の各種リソースへの参照指示を示す。RDA Registry への参照も、これに含める。
- ・「:referredVES」RDA の値語彙である VES (Vocabulary Encoding Scheme) への参照指示を示す。

e) 例示の記録

個別規定には、当該規定に従い記録された値が例示として含まれる場合がある。例示はその値（文字列）のみが示されるものと、補足情報がそれに続くものとがある。RDF 化では、こうした組み合わせをそのまま表現するよう、プロパティ「:example」と空白ノードによる構造化表現を用いる。例示の値はプロパティ「rdf:value」の目的語に位置づけ、それ以外の補足部分はプロパティ「:note」を用いる。

なお、「View in Context Example」の下に示される例示については、複数の値とそれぞれの補足情報をまとめて単一の事例が構成されているため、さらなる構造化表現を採用した。

「View as Relationship Example」という例示についても、それに即した構造化表現とした。

3. RDF データへの変換処理

上述した検討結果を踏まえ、RDA のインストラクションとガイダンスに対して RDF 化を試行した。今回は、下記を対象とした。

- ・ガイダンス (34)
- ・実体に関するインストラクション (13)

- ・体現形のエレメントに関するインストラクション (400)
- ・著作のエレメントに関するインストラクション (590)

対象としたエレメントは、RDA 語彙である RDA-Vocabularies-4.1.2 (2022 年 1 月) に登録されていた体現形と著作のすべてのエレメントとした（「deprecated」とされていたものは除外）。

変換作業の手順は次のとおりである。RDA Toolkit 上で、対象としたガイダンス／インストラクションのウェブページを表示させた後、HTML ファイルとしてローカルに保存した。それらファイルからプログラムによるスクレイピングによって RDF データに機械的に変換し生成した。その後、筆者らが手作業で確認と修正を加えた。

RDF データ化した RDA 規定の集計結果を第 1 表および第 2 表に示す。表内では、「実体」とは「実体インストラクション」を、「体現形」とは「著作」とはそれぞれのエレメントのインストラクションを指す。今回の変換処理で重大な問題は見当たらなかったことから、本研究の提案方式には一定程度の妥当性があると考ええる。

4. Policy statements の変換処理

レイア 3 に位置づけた適用細則として、RDA Toolkit 上では現在 3 種の Policy statements が公開されている。具体的には、米国議会図書館と PCC 参加館による LC-PCC PS、英国図書館の BLPS、そして米国音楽図書館協議会のベストプラクティス MLA BP である。これらに対応する RDA のインストラクション等に依拠した形式で提示されており、RDF データ化もそれに即した方針とした。

URI の付与単位は独立して扱える最小単位とし、今回は便宜的に参照先 RDA 規定の URI をベースにしたものとする。RDA 規定との対応づけにはプロパティ「:policyFor」、指示内容文にはプロパティ「:policy」をそれぞれ用いて表現した。下記に LC-PCC PS の一例を示す。

```
rdam:P30156iLCPS¥#03.01.02
rdf:type :RDA-LC-PCC-policyStatement ;
rdfs:label "Policy Statement: title proper --
Recording -- Recording an unstructured
description" ;
:instrectFor rdam:P30156 ;
:policyFor rdam:P30156i¥#03.01.02 ;
:policy "LC/PCC practice: Apply the option
for serials. Apply the option for other ...";
```

:policy "If the manifestation has only a single title and the title appears on ...";
 dcterms:modified "2021-07-20".

また、個別規定からメタデータへの参照、メタデータから適用された個別規定への参照の問題については、NCR2018 と同じ構図が当て

はまることを確認した。

引用文献

- 1) 谷口祥一, 橋詰秋子. 記述規則 NCR2018 の RDF データ化. *Library and Information Science*. 2022, no. 87 (印刷中).

```
rdam:P30156i¥#03.03.01
rdf:type :RdaInstruction ;
rdfs:label "title proper -- Recording -- Titles proper of manifestations of parts and iterations" ;
:instretFor rdam:P30156 ;
:condition "A manifestation is a part or iteration of a manifestation." ;
:condition "A value of Manifestation: title of manifestation includes a common title." ;
:referredInstret rdam:P30134i # title of manifestation;
:option "Record a value that includes a common title." ;
:isPartOf rdam:P30156i¥#03.03 ;
dcterms:modified "2019-04-30".
```

第1図 RDF データ例：エレメント「title proper」のインストラクション（一部）

第1表 プロパティの出現回数

	ガイダンス 34件		実体 13件		体現形 400件		著作 590件	
	出現数	平均	出現数	平均	出現数	平均	出現数	平均
規定数 (URI 付与数)	514	15.12	235	18.08	6,137	15.34	5,720	9.69
:instruction	703	20.68	293	22.54	4,206	10.52	6,633	11.24
:option	241	7.09	87	6.69	2,169	5.42	326	0.55
:condition	188	5.53	68	5.23	996	2.49	146	0.25
:example	49	1.44	10	0.77	153	0.38	84	0.14
:referredInstret	328	9.65	208	16.00	3,006	7.51	4,552	7.72
:referredEntyInstret	63	1.85	79	6.08	66	0.17	45	0.08
:referredGuidance	123	3.62	100	7.69	1,926	4.82	2,376	4.03
:referredVES	14	0.41	0	0.00	81	0.20	9	0.02
:referredResource	67	1.97	0	0.00	2	0.01	0	0.00
:hasSection, :isSectionOf	223	6.56	141	10.85	3,455	8.64	4,790	8.12
:hasPart, :isPartOf	257	7.56	81	6.23	2,282	5.71	340	0.58

第2表 個別規定の集計

	ガイダンス		実体		体現形		著作	
	規定数	平均	規定数	平均	規定数	平均	規定数	平均
指示内容文をもつ規定	418	12.29	179	13.77	4,417	11.04	3,868	6.56
:instruction のみ	177	5.21	105	8.08	2,469	6.17	3,584	6.07
:option のみ	95	2.79	24	1.85	1,015	2.54	152	0.26
:condition+:option	146	4.29	50	3.85	852	2.13	126	0.21
:instruction+:option	0	0.00	0	0.00	81	0.20	6	0.01
指示内容をもたない規定	96	2.82	56	4.31	1,720	4.30	1,852	3.14
参照指示をもつ規定	243	7.15	152	11.69	3,028	7.57	3,652	6.19
:referredInstret をもつ	162	4.76	97	7.46	1,965	4.91	2,915	4.94
:referredEntyInstret をもつ	24	0.71	56	4.31	66	0.17	45	0.08
:referredGuidance をもつ	75	2.21	92	7.08	1,844	4.61	2,376	4.03
参照指示をもたない規定	271	7.97	83	6.38	3,109	7.77	2,068	3.51

注：平均は、ガイダンス単位、実体インストラクション単位、またはエレメントインストラクション単位