

NCR2018 と RDA による語彙の RDF 定義とメタデータスキーマ

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

NCR2018 と RDA (Toolkit ベータ版に対応) による語彙の RDF 定義に対して、①概念モデルを適用した段階に対応する定義、②属性および関連の要素の「記録の方法」の指示に依拠した定義、③関連の要素を再構成した定義、④上記以外の構造表現力を導入した定義という、4層からなる理解と評価を目的にした枠組みを提示し検討した。併せて、②から④の各層において有効な代替案または拡張等を、NCR2018 と RDA それぞれについて提示した。

1. はじめに

NCR2018 については、JLA 目録委員会のウェブページ「エレメント・語彙等データ提供」¹⁾において、採用した実体、エレメント、語彙リストの用語、関連指示子について、Excel と csv 形式という機械可読データとして定義情報が提供されている。これはリンクトデータ化に向けた最初のステップと位置づけられる。今後、必要な RDF クラスとプロパティによる定義を策定し公開していくこと、およびそれらに基づき適切なメタデータスキーマを構成し、メタデータの作成と提供を図ることが求められる。

一方、RDA は以前からメタデータレジストリ RDA Registry²⁾において、その語彙の RDF 定義情報を公開してきている。2019 年には、依拠する概念モデルを IFLA LRM に変更するなど、各種の変更を組み入れた 3R プロジェクトを完了し、RDA Toolkit においてベータ版として変更後の RDA を公開するとともに、対応する語彙の RDF 定義を上記の RDA Registry においていち早く公開している。2020 年末にはベータ版が正式版に移行する予定である。ただし、変更後の RDA エレメント、そしてその語彙定義は複雑な様相を呈しており、容易には理解しにくい、使いこなせないものに見える。

本研究は、NCR2018 の語彙に対して、どのような枠組みに沿って RDF 定義を策定すべきか、複数の可能な RDF 定義をどのように位置づけるべきかなど、理解と評価に向けた枠組みを提案し検討する。また、公開済み RDA 語彙に対して、同様の観点から、提案した枠組みに沿って評価と代替案の提示を試みる。

提案する枠組みとは、レイア 1: 概念モデル (実体関連モデル) とそれを NCR2018/RDA に適用した段階において導かれる RDF 定義、レイア 2: 属性および関連の要素ごとの「記録の方法」として指示された方法に依拠した RDF 定義、レイア 3: 関連の要素を再構成した RDF 定義、そしてレイア 4: 上記

以外の構造表現力を導入した RDF 定義、という 4層から構成する。

2. NCR2018 語彙の RDF 定義

2.1 レイア 1: 概念モデルに対応する語彙定義

NCR2018 は、実体関連モデルによる FRBR を概念モデルとしており、実体、関連、属性から構成される。なお、FRBR では、関連に属性は設定していない。

NCR2018 を表す概念モデルは、11 の実体 (第 1 グループ: 著作、表現形、体現形、個別資料、第 2 グループ: 個人、家族、団体、第 3 グループ: 概念、物、出来事、場所) からなる。実体にはそれぞれ属性が設けられており、それらは NCR において「エレメント」として規定されている。実体間の関連も同様にエレメントとされ、加えて関連の詳細な種類を示す「関連指示子」もエレメントに準じて位置づけられる。

さらに、概念モデルでは表現されない値リストの語彙を加えて、これらの定義情報を Excel と csv ファイルで公開している。これらは、概念モデルとそれに対応する語彙の理解を目的とすると位置づけられる。

公開されている項目群は下記の通りである。

ID、種別、条項番号、名称、名称 (英語)、定義域、値域、対応エレメント、上位、逆方向、定義、更新日、備考

「種別」は、「実体、エレメント、サブエレメント、エレメント・サブタイプ等」の区分を示す。「定義域」は属性のエレメントの場合、当該エレメントが属する実体、関連のエレメントおよび関連指示子の場合、関連元となる実体を表し、他方、「値域」は関連のエレメントおよび関連指示子の場合、関連先となる実体を表す。また、「対応エレメント」とは、語彙のリストの用語に対して、当該用語が用いられるエレメント、関連指示子に対して、当該指示子が用いられる関連のエレメントを指す。

次に、これら語彙定義を RDF モデルに変換し、RDF クラスおよびプロパティによって表現する。概念モデルの実体は、URI を割り当てて RDF クラスとする。たとえば、クラス「著作」は <http://jla.or.jp/data/ncr2018/C100001> とする。加えて、モデル構成上、便宜的に新規クラス「NCR2018 実体 (NCR2018 entity)」（NCR2018 で定義されているすべての実体の上位クラス）、「書誌的実体 (Bibliographic entity)」（著作から個別資料までの 4 実体 WEMI の上位クラス）、「行為主体 (Agent)」（個人、家族、団体の上位クラス）を導入する。

実体の属性は、すべて RDF プロパティとなり、当該実体がプロパティの定義域となる。また、エレメントとエレメント・サブタイプ（例：タイトルと本タイトル、並列タイトル等）は、上位-下位のプロパティとなるが、エレメントとサブエレメントの場合（例：出版表示と出版地、出版者等）にはこうした関係にはない。なお、いずれの場合も、概念モデルに依拠して値域の指定はない。

一例として、エレメント・サブタイプである「本タイトル」について RDF 定義例を示す。

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ncr: <http://jla.or.jp/data/ncr2018/> .
@prefix vcbncr: <http://jla.or.jp/vocab/ncr2018#> .
ncr:E200002
  vcbncr:id "200002" ;
  rdf:type rdf:Property ;
  vcbncr:type "22" ;
  vcbncr:typeLabel "エレメント・サブタイプ" ;
  vcbncr:instructionNumber "#2.1.1" ;
  rdfs:label "本タイトル" ;
  rdfs:label "title proper"@en ;
  <http://www.w3.org/2002/07/owl#equivalentProperty>
    <http://rdaregistry.info/Elements/m/P30156> ;
  rdfs:domain ncr:C100003 ;
  rdfs:subPropertyOf ncr:E200001 ;
  <http://www.w3.org/2004/02/skos/core#definition>
    "表現形を識別するための固有の名称。" ;
  <http://purl.org/dc/terms/modified>
    "2019-03-22" .
```

なお、上記の定義域は「表現形」、上位プロパティは「タイトル」を表す。

実体間の関連（と関連指示子）も、すべて RDF プロパティとなるが、関連元と関連先の実体がそれぞれ定義域と値域に指定される。たとえば、プロパティ「表現形から著作への関連」では、定義域が表現形、値域が著作となる。

また、語彙リストの用語（機器種別「オーデ

イオ」など）は、RDF クラス（ただし、`skos:Concept`）として定義される。併せて、統制形アクセス・ポイント等も、プロパティとしてエレメントと同様に定義できる。

これらの RDF 定義は、あくまでも概念モデルに対応した RDF モデルとその語彙定義を理解することを主目的にすると位置づけられる。

2.2 レイア 2: 記録の方法に対応させた語彙定義

NCR2018 は、個々のエレメントに対して、「記録の方法」を定めており、これは記録する値の形式を指定していることになる。実体の属性に対応するエレメントは、その値の形式等について元々限定がないため、RDF 定義への変換においてプロパティは値域の指定をもたない定義とされる。それゆえ、いかなる記録の方法が指定されても齟齬は生じない。

一方、実体間関連に対応するエレメントについては、その記録の方法として、識別子、典拠形アクセス・ポイント（典拠形 AP）、複合記述、構造記述、非構造記述のいずれか（関連の種類によって限定あり）を NCR は指定している。これらを RDF 定義の観点から見ると、識別子は URI に組み替えることが通常可能であるが、それ以外は典拠形 AP を含めて、リテラルである。この点から、概念モデルに対応する RDF 語彙定義において値域が指定されているプロパティは、それが整合しないことになる。たとえば、プロパティ `ncr:E200335` 「表現形から著作への関連」の値域は著作と指定されているが、その記録の方法では識別子、典拠形 AP、そして複合記述が認められており、そのため値としてリテラルを許容する必要がある。こうした「記録の方法」に合致するよう RDF 定義を変更するには、以下の 2 つの方法が考えられる。

方法 1: 値域の指定をもつプロパティ（オブジェクト・プロパティ）ではなく、値域の指定をもたないプロパティに変更する。こうしたプロパティを、RDA 語彙定義では `canonical` プロパティと呼んでいる。そのため正式には、元のオブジェクト・プロパティをそのまま残し（あるいは、値域をリテラルとするデータタイプ・プロパティを加えた上で）、その上位プロパティとして新たに `canonical` プロパティを導入する。そして、これを先ほどの関連のエレメントに対応づける。これは RDA において採用されている方法である。

方法 2: 値域の指定を、URI とリテラルの両者を包含したものに変更する。そのためには、a)

元のプロパティにおいて、値域を下記のように指定する (ncr:C100001 は「著作」)。

```
ncr:E200335
  schema:rangeIncludes
    ncr:C100001, rdfs:Literal .
```

schema:rangeIncludes は、schema.org が便宜的に複数クラスの和集合を値域とするために導入したプロパティである。

あるいは、b) 関連先が著作であるエレメントに対して、新たなクラス WorkOrItsDescription (著作またはその記述) を値域に指定し、この新クラスを下記のように定義する。

```
ncr:WorkOrItsDescription
  rdf:type owl:Class ;
  owl:unionOf (ncr:C100001 rdfs:Literal) .
```

同様に、関連先が行為主体であるエレメントに対して、クラス AgentOrItsAppellation (行為主体またはその名称) を値域に指定し、このクラスを上記と同様に定義する。

こうして変更を加えた RDF 定義は、メタデータ公開・提供用のスキーマを構成するものと位置づけることが適切と考える。

2.3 レイア 4: 追加的な構造表現力を導入した語彙定義

エレメントの下でのエレメント・サブタイプのグループ化やサブエレメントのグループ化が、メタデータ作成用スキーマとして有効と考える。特に単一の記述において複数回出現が認められるエレメント (出版表示、シリーズ表示など) には、個別の出現を相互に区別するために不可欠である。むしろ、こうしたグループ化は、メタデータ公開・提供用スキーマにおいても適用することが可能である。

このグループ化のためには、エレメントに対応するプロパティの値域に新規クラス (例: PublicationStatement (出版表示クラス)) を設け、サブエレメント等に対応するプロパティの定義域にこの新規クラスを指定する。ある事例に当てはめたときには、下記のようになる。

```
ex:体現形インスタンス 1 ncr:E200032 [
  rdf:type ncr:PublicationStatement ;
  ncr:E200033 "出版地 1" ;
  ncr:E200035 "出版者 1" ;
  ncr:E200037 "出版日付 1" ] .
```

他方、値自体の構造化した表現が、日本語表示形に対する読みの付与の場合も含めて必要である。そのためには、RDA に倣い、クラス Nomen とそれを定義域とするプロパティ「nomen 文字列」「文字種」「言語」などの導

入が有効であろう。

```
ex:体現形インスタンス 1 ncr:E200002 [
  rdf:type ncr:Nomen ;
  ncr:nomen 文字列 "タイトル 1" ;
  ncr:読み "読み 1" ;
  ncr:言語 "日本語" ] .
```

3. RDA 語彙の RDF 定義

3.1 レイア 1: 概念モデルに対応する語彙定義

RDA は依拠する概念モデルを IFLA LRM に切り替えており、その結果、実体 Nomen、Place、Timespan、Agent 等が導入され、これまでの WEMI 等とすべての組み合わせ (13 実体×13 実体) が実体間関連として発生している。これらすべてが RDA エレメントになるとともに、個々の関連が詳細に展開されている。また、個別実体ごとの多数の属性、さらには関連指示子とされてきたものもすべて RDA エレメントとなり、そのまま RDF によるプロパティとして定義され、結果的に膨大な数のプロパティの集積となって提供されている。

たとえば、以前は体現形の属性であったエレメント "place of publication" や "place of manufacture" 等とその値の記録は、a) 体現形から Place への関連、b) Place から Nomen への関連、そして c) Nomen の属性である nomen string による組み合わせをもって記録されることになる。

RDF 定義では、a) のプロパティ rdam:P30088 ("has place of publication") は概念モデルに依拠すれば定義域が rdac:C10007 (体現形)、値域が rdac:C10009 (Place) となる。さらに、その上位プロパティには rdam:P30279 ("has place of manifestation")、さらに rdam:P30272 ("has related place of manifestation") などがある。他方、下位には rdam:P30092 ("has parallel place of publication") が設定されている。

次に、Place から Nomen への関連に対応して、プロパティ rdap:P70012 ("has related nomen of place")、その下位プロパティ rdap:P70019 ("has appellation of place")、さらに下位の rdap:P70001 ("has name of place")、rdap:P70018 ("has access point for place")、rdap:P70020 ("has identifier for place") などを定義している。これらは、概念モデルに従えば、いずれも定義域が rdac:C10009 (Place)、そして値域が rdac:C10012 (Nomen) となる。

このように、名称に相当する実体 **Nomen** を導入したことにより、**Nomen** 以外の実体は、その名称を表す属性をもつことなく、**Nomen** への関連をもって表すことになる。

そして最後に、**Nomen** を定義域とするプロパティ `rdan:P80068 ("has nomen string")` をもって、リテラルである名称などを記録することになる。このときに `rdan:P80066 ("has language of nomen")` など、**Nomen** の属性に該当するプロパティを併せて適用できる。

3.2 レイア 2: 記録の方法に対応させた語彙定義

RDA は個々のエレメントに対して、「記録の方法」を定めており、非構造記述、構造記述、識別子、そして IRI (URI) の 4 つとしている。よって、RDA では、URI 以外はリテラルの扱いとなる。たとえば、エレメント **place of publication** については、非構造記述では `name of place` の値を、構造記述では `access point for place` の値を、識別子では `identifier for place` の値を、URI では対象とする **Place** がもつ URI を、それぞれ記録すると規定している。つまり、**Place** の URI、または名称などリテラルを記録するという指示である。これは、概念モデルに依拠した RDF 定義とは異なる指示となる。

こうしたエレメントの「記録の方法」に合致するよう、公開済みの RDF 定義では、個々のエレメントに対応するプロパティとして最大 4 つ、a) 値域を指定しない canonical プロパティ (`rdam:P30088 ("has place of publication")`)、b) 値域に特定クラスを指定したオブジェクト・プロパティ (`rdam/object:P30088`)、c) 値域をリテラルとするデータタイプ・プロパティ (`rdam/datatype:P30088`)、そして d) 定義域と値域のいずれも指定しない非制約プロパティ (`rdau:P60163 ("has place of publication")`) を含めている。そして、b) と c) の上位が a)、さらにその上位が d) となる。

これらによって、記録の方法に対応できることになるが、元々膨大な数のプロパティに対して、さらにその数を倍増させる結果をもたらしている。

そこで、RDA に対しても、NCR2018 について示した「方法 2」の a) または b) を適用することを提案する。

3.3 レイア 3: 関連のエレメントを再構成した語彙定義

公開済み RDA 語彙は、膨大な数のプロパテ

ィを宣言しており、また関連のエレメントに対応する多くのプロパティは多段階の階層関係にある。これは語彙の理解そして使い勝手を悪くしている。

こうした RDF 定義に対して、以下のような代替案を提案する。限定された数の汎用プロパティと詳細なクラス階層を用いて、現行の語彙と同じ意味を表現することを意図している。たとえば、体現形から **Place** への関連については、下記のような RDF 表現とする。

```
rdac:C10007 rdam:P30272 [  
  rdf:type :PlaceOfPublication ;  
  rdf:value :PlaceOrItsAppellation ] .
```

ここで `rdam:P30272` は "has related place of manifestation" であり、クラス `:PlaceOrItsAppellation` は、先の「記録の方法」への対応の箇所ですべて「方法 2」によって仮に定義したクラス「**Place** またはその名称」であり、代えてクラス `Place` でも、あるいは値域を指定しない canonical プロパティとしてもよい。併せて、`:PlaceOfPublication` については、下記のように必要な詳細化と階層化を図る。

```
:PlaceOfPublication  
  rdfs:subClassOf :PlaceOfManifestation .  
:PlaceOfDistribution  
  rdfs:subClassOf :PlaceOfManifestation .  
:PlaceOfManifestation  
  rdfs:subClassOf rdac:C10009 .
```

具体的な事例を記述すると、下記ようになる。

```
:ex:Manifestation1 rdam:P30272  
  [ rdf:type :PlaceOfPublication ;  
    rdf:value ex:Place1 ] ,  
  [ rdf:type :PlaceOfDistribution ;  
    rdf:value "place2" ] .
```

これらの結果、プロパティの数が限定され、かつ比較的大きな粒度での設定となり、他方その意味の詳細についてはクラスの階層で表現する。こうした変更は、メタデータ公開・提供用スキーマを構成する上での代替案と考える。なお、これは概念モデルにおいて実体間関連に属性を認めたときの対応する RDF 表現に相当する。

また、レイア 4: 追加的な構造表現力を導入した RDA 語彙定義を、NCR2018 の場合と同様に想定することができる。

引用文献

- 1) 日本図書館協会目録委員会. NCR2018 年版エレメント・語彙等データ提供. 2019.
<https://www.jla.or.jp/Portals/0/data/iinkai/mokuroku/ncr2018//tabid/795/Default.aspx>
- 2) RDA Registry. <http://www.rdaregistry.info/>