

RDF オントロジーのマッピングとマージの形式化

谷口 祥一 (慶應義塾大学文学部)

taniguchi@z2.keio.jp

1. はじめに

現在、多数かつ多様な RDF データが Linked Data として公開・提供されている。また、それらは領域においても多岐にわたり、複数領域の横断的なものから特定領域に限定されるものまで幅がある。こうした RDF データには、1)対象とする個別のリソース (インスタンス) に対するメタデータ記述と、2)そうしたメタデータ記述に用いる語彙 (RDF クラスとプロパティの定義情報) 自体を RDF データとして公開したもの、あるいは 3)データ値の記録に用いる値を RDF クラスとして定義し公開したものなどが含まれる。図書館等に直接かかわるものとしても、1)情報資源とそれらに関するエージェント等に関するメタデータ、2)それらメタデータ記述に用いるエレメントなどの定義情報、3)主に情報資源の特定のエレメント記述または主題表現に用いる概念体系 (RDF クラス群) などがある。上記 2 および 3 は、オントロジーとも呼ばれる。

本研究では、このうち上記 2、および 3 の一部に属するものを対象とし、RDF と RDFS が規定するクラスとプロパティのみを用いて定義されているオントロジーを想定し、複数の異なるオントロジー間のマッピングとマージの形式化、すなわちそれらが成立する要件の明示およびマージにおける選択肢の明示を試みる。これらは、そのままマッピングとマージの手順の提示に通じる。

発表者は以前に IFLA Library Reference Model (LRM) と BIBFRAME という 2 つの語彙のマッピングとマージを試行しており¹⁾²⁾、その過程で必要とされた要件の明示等を、本研究は改めて試みる。

なお、オントロジーのマッピングやマージに関する研究は相当数の蓄積があるが³⁾⁵⁾、管見の範囲では、本研究が取り上げる事項について詳細に検討したものは見受けられない。

2. 対象とするオントロジー

RDF と RDFS が規定するクラスとプロパティのみを用いて定義されている語彙に対象を限定する。具体的には、以下のクラスとプロパティを適用して定義された語彙とする。

`rdfs:Resource`, `rdfs:Class`, `rdfs:Literal`,

`rdf:Property`, `rdfs:Datatype`, `rdf:type`,
`rdfs:subClassOf`, `rdfs:subPropertyOf`,
`rdfs:domain`, `rdfs:range`, `rdfs:label`,
`rdfs:comment`, `rdfs:seeAlso`,
`rdfs:isDefinedBy`

なお、`rdf:type` はインスタンスを示すプロパティであるが、RDF/RDFS ではあるリソースがクラスかつ他のクラスのインスタンスであることは認められており、定義情報において `rdf:type` は出現しうる。また、`rdfs:seeAlso`, `rdfs:isDefinedBy` など、本研究の範囲では直接参照しないプロパティも含めている。

他方、以下の RDF/RDFS クラスとプロパティは対象外とする。

`rdf:Statement`, `rdfs:Container`, `rdf:Bag`,
`rdf:Seq`, `rdf:Alt`, `rdf>List`, `rdf:first`, `rdf:rest`,
`rdf:subject`, `rdf:predicate`, `rdf:object`

これらは、RDF トリプルのメタ記述 (具体化、*reification*) に用いるものと、コンテナ/コレクションの記述に用いるものであり、通常、オントロジーには出現しない。

また、SKOS や OWL のクラスとプロパティを用いて定義された語彙は対象外とする。ただし、`skos:definition` がクラスまたはプロパティの意味範囲の定義を示すために使用されているときには、そのみ採用する。

こうしたオントロジーとしては、書誌的世界に関するものとして FRBR, FRAD, FRISAD, IFLA LRM, BIBFRAME, FRBRoo, schema.org などがある。なお、ER モデルで表現されている FRBR などは、RDF モデルに変換 (実体は RDF クラスに、属性と関連は RDF プロパティに変換) した上で対象となる。

本研究では、クラスやプロパティの意味定義 (意味範囲) の包含関係が機械的に判断できない状況を想定しており、それゆえ人手によるマッピングとマージの処理となる。

3. RDF オントロジーのマッピング

2 つのオントロジーの間でのクラス間、プロパティ間の、方向性をもった対応づけがマッピング (*mapping*) である。マッピングにおいて、クラス定義やプロパティ定義の変更や追加はない。また、クラス間のサブクラス関係やプロパティ間のサブプロパティ関係の追加もない。マッピングは、1 対 1 または 1 対多の対応づけ

となるが、対応づけがなされない、すなわちマッピング先がないクラスやプロパティは発生しうる。また、逆向きのマッピングはそれ自体を独立したマッピングとして規定する必要がある。

3. 1 クラスのマッピング

オントロジー O_1 のクラス C_i に対して、マッピングとは、オントロジー O_2 のクラスのうち、下記の条件①および②を満たす(① \wedge ②)クラス C_j を同定すること、そして同定されたクラスがマッピング結果(マッピング先)となる。

【条件①】クラス C_i が意味範囲においてクラス C_j に包含される。

述語論理記法にRDF/Turtleの記法を組み合わせると、インスタンス $:x$ に対して、下記の条件となる。

$$\forall x (:x \text{ rdf:type } C_i . \rightarrow :x \text{ rdf:type } C_j .)$$

実際には、意味範囲の定義はプロパティ `rdfs:comment` (または `skos:definition`) を用いて自然言語で記述されるため、それを解釈し包含関係を判断することになる。

【条件②】 C_j は上記の条件①を満たす、意味範囲において最下位レベルのクラスとする。すなわち、複数のマッピング先候補があるときには、そのうち最下位レベルのクラスを選択する。

C_j はオントロジー O_2 に属する、 C_j 以外のクラスとすると、上記の条件は下記を満たす C_j を求めることを指す。

$$\forall x' (:x' \text{ rdf:type } C_j . \rightarrow :x' \text{ rdf:type } C_j .) \wedge \neg \forall x' (:x' \text{ rdf:type } C_j . \rightarrow :x' \text{ rdf:type } C_j .)$$

すなわち、 C_j のインスタンスであって、かつ C_j のインスタンスでない x' が存在する。

上記の条件①および②を満たさない場合には、クラス C_i のマッピングはなされない。たとえ意味的にクラス間に部分的な包含関係、すなわち重なりがあるときをも、これは含む。

また、当該クラスを定義域とするプロパティを含めてクラスの意味定義とする扱いはここでは採用しない。プロパティのクラスからの独立性がRDFでは強いいため、両者を切り分けて扱う。

3. 2 プロパティのマッピング

オントロジー O_1 のプロパティ P_i に対して、マッピングは下記の条件①から④のすべてを満たす(① \wedge ② \wedge ③ \wedge ④)オントロジー O_2 のプロパティ P_j を同定すること、同定されたものがマッピングの結果(マッピング先)となる。

【条件①】プロパティ P_i が意味範囲において

プロパティ P_j に包含される。

ここで、プロパティ P_i の定義域を D_i 、プロパティ P_j の定義域を D_j とする。すなわち、

$$P_i \text{ rdfs:domain } D_i . \quad P_j \text{ rdfs:domain } D_j .$$

同様に、プロパティ P_i の値域を R_i 、プロパティ P_j の値域を R_j とする。すなわち、

$$P_i \text{ rdfs:range } R_i . \quad P_j \text{ rdfs:range } R_j .$$

上記の条件①を述語論理記法で表すと、

$$\forall x, y (:x \text{ rdf:type } D_i . \wedge :y \text{ rdf:type } R_i .$$

$$\wedge :x \text{ } P_i \text{ } :y . \rightarrow :x \text{ } P_j \text{ } :y .)$$

プロパティにはクラスとそのインスタンスのような区別はないため、上記の表現で代用することになる。上記の論理式は、実際にはサブプロパティを定義するものであり、そこに定義域と値域の条件指定が部分的にせよ、含まれてしまっている。

なお、プロパティの意味範囲の定義は、通常、`rdfs:comment` (または `skos:definition`) に記述されるため、それを解釈し判断することになる。

【条件②】プロパティ P_i の定義域 D_i 、プロパティ P_j の定義域 D_j に関して、下記のa)またはb)が成立する(a \vee b)。

a) 定義域として指定されたクラス D_i が意味範囲においてクラス D_j に包含される。

$$\forall x (:x \text{ rdf:type } D_i . \rightarrow :x \text{ rdf:type } D_j .)$$

b) 定義域 D_j が未指定(unspecified)。これは下記と同等である。

$$P_j \text{ rdfs:domain rdfs:Resource .}$$

つまり、

$$\forall x (:x \text{ rdf:type rdfs:Resource .}$$

$$\rightarrow :x \text{ rdf:type } D_j .)$$

`rdfs:Resource`は`rdfs:Class`と`rdfs:Literal`を含んでおり、未指定の場合と同等となる。

なお、このとき定義域 D_i は、クラス指定であっても、未指定であってもよい。

【条件③】プロパティ P_i の値域 R_i 、プロパティ P_j の値域 R_j に関して、下記のa)、b)、c)のいずれかが成立する(a \vee b \vee c)。

a) 値域として指定されたクラス R_i が意味範囲においてクラス R_j に包含される。

$$\forall y (:y \text{ rdf:type } R_i . \rightarrow :y \text{ rdf:type } R_j .)$$

b) 値域 R_i と値域 R_j がいずれもリテラル。

$$P_i \text{ rdfs:range rdfs:Literal .}$$

$$P_j \text{ rdfs:range rdfs:Literal .}$$

$$\forall y (:y \text{ rdf:type rdfs:Literal .}$$

$$\rightarrow :y \text{ rdf:type } R_j . \wedge :y \text{ rdf:type } R_j .)$$

c) 値域 R_j が未指定(unspecified)。

$$P_j \text{ rdfs:range rdfs:Resource .}$$

$$\forall y (:y \text{ rdf:type rdfs:Resource .}$$

$$\rightarrow :y \text{ rdf:type } R_j .)$$

このとき値域 R_i は、クラスであっても、未指定であってもよい。

【条件④】プロパティ P_j は上記①～③の条件を満たす、意味範囲において最下位レベルのプロパティとする。これは複数のマッピング先候補が存在するときには、最下位レベルのプロパティを選択することを指す。

P_j はオントロジー O_2 に属する、 P_j 以外のクラスとすると、下記を満たす最下位クラス P_j を求めることになる。

$$\begin{aligned} & \forall :x', :y' (:x' \text{rdf:type } D_j. \wedge :y' \text{rdf:type } R_j. \\ & \quad \wedge :x' P_j :y'. \rightarrow :x' P_j :y'.) \wedge \\ & \neg \forall :x', :y' (:x' \text{rdf:type } D_j. \wedge :y' \text{rdf:type } R_j. \\ & \quad \wedge :x' P_j :y'. \rightarrow :x' P_j :y'.) \end{aligned}$$

4. RDF オントロジーのマージ

オントロジー O_1 と O_2 から、新たな O_3 を導き出すのがマージ (merge) である。マージ結果のオントロジー O_3 は O_1 と O_2 のクラス、プロパティの全部またはその一部を含む。それゆえ、元の O_1 と O_2 のクラスとプロパティをすべて取り込むことを求めるものではない。

マージ処理のレベルを、本研究では、下記の3つに区分する。

レベル1: O_1 と O_2 のクラス、プロパティの採用と関係づけ。 O_1 と O_2 の既存クラス、プロパティに修正を加えず、 O_3 においてそれらをそのまま採用する。

レベル2: O_1 と O_2 のクラス、プロパティの調整 (修正)。 O_3 において整合するよう、「最小限の」調整を加える。あるいは、調整結果に相当する新規のクラス、プロパティを追加する。

レベル3: 上記の範囲を超えたクラス、プロパティの修正と追加。 O_1 と O_2 の各クラスやプロパティの上位クラスや上位プロパティを新たに定義し、既存のものをその下位に位置づけることなども含まれる。例えば、FRBRoo は、FRBR 等と CIDOC CRM を対象にしてレベル3に該当するマージを行った例と捉えることができる。同様に、IFLA LRM は、FRBR, FRAD, FRISAD の単純なマージを超えてレベル3に相当するマージを実行した例である。

4. 1 クラスのマージ：レベル1

1) オントロジー O_1 のクラス C_i , オントロジー O_2 のクラス C_j に関して、クラス C_i と C_j が意味範囲において重なりがないとき (すなわち、クラス C_i がすべての C_j と意味範囲において重なりがなく、かつクラス C_j がすべての C_i と意味範囲において重なりがないとき) は、下記

のように表すことができ、

$\neg \exists :x (:x \text{rdf:type } C_i. \wedge :x \text{rdf:type } C_j.)$
オントロジー O_3 のクラスとして C_i と C_j の両者を採用する。

2) 他方、クラス C_i が意味範囲においてクラス C_j と重なりがあるときは、下記のように表すことができ、

$\exists :x (:x \text{rdf:type } C_i. \wedge :x \text{rdf:type } C_j.)$
マージ処理は、下記の選択肢 1 から 3 のいずれかとなる。

- ・選択肢 1: オントロジー O_1 のクラス C_i とオントロジー O_2 のクラス C_j の両者を採用する。
- ・選択肢 2: クラス C_i を採用する。
- ・選択肢 3: クラス C_j を採用する。

いずれを選択するのかは任意であり、それゆえ通常は依拠する方針に基づく選択となる。

加えて、上記の選択肢 1~3 において、オントロジー O_3 におけるその時点のクラス C_k (オントロジー O_3 は順次更新される) と採用したクラス C_i または C_j が意味範囲において包含関係があるときには、すなわち

$$\forall :x (:x \text{rdf:type } C_i. \rightarrow :x \text{rdf:type } C_k.)$$

または、

$$\forall :x (:x \text{rdf:type } C_j. \rightarrow :x \text{rdf:type } C_k.)$$

と表され、マージ処理において、下記のクラス間の関係づけを追加する。

$$C_i \text{ rdfs:subClassOf } C_k.$$

または、 $C_j \text{ rdfs:subClassOf } C_k.$

なお、 C_k は上記の条件を満たす、意味範囲において最下位レベルのクラスとする。

4. 2 クラスのマージ：レベル2

クラス C_i が意味範囲においてクラス C_j と重なりがあるとき、マージ処理は、下記の選択肢 1 から 3 のいずれかとなる。

- ・選択肢 1: オントロジー O_1 のクラス C_i とオントロジー O_2 のクラス C_j の両者を等しく調整し採用する。
- ・選択肢 2: クラス C_i を優先して採用し、それと整合するようクラス C_j を調整し採用する。
- ・選択肢 3: クラス C_j を優先して採用し、それと整合するようクラス C_i を調整し採用する。

なお、調整とは、クラスのマッピングにおいて示した条件に合致させる修正を指す。調整 (修正) ではなく、調整結果に相当する新たなクラスを追加するという選択肢もある。

加えて、上記の選択肢 1~3 において、オントロジー O_3 におけるその時点のクラス C_k (オントロジー O_3 は順次更新される) と採用したクラスが意味範囲において包含関係があると

きには、レベル1のマージ処理と同様、下記のクラス間の関係づけを追加する。

$C_j \text{ rdfs:subClassOf } C_k$.

または、 $C_j \text{ rdfs:subClassOf } C_k$.

ただし、 C_k は上記の条件を満たす、意味範囲において最下位レベルのクラスとする。

4.3 プロパティのマージ：レベル1

あくまでも、クラスのマージにおけるレベル1に対応して採用するものである。これは、プロパティの定義域、値域に指定されているクラス自体の採用が前提となるためである。

1) オントロジー O_1 のプロパティ P_i とオントロジー O_2 のプロパティ P_j に関して、プロパティ P_i と P_j が意味範囲において重なりがないとき（すなわち、プロパティ P_i がすべての P_j と意味範囲において重なりがなく、かつプロパティ P_j がすべての P_i と意味範囲において重なりがないとき）には、下記のように表すことができ、

$$\neg \exists x, y (:x \text{ rdf:type } D_i. \wedge :y \text{ rdf:type } R_j. \\ \wedge :x \text{ } P_i :y. \wedge :x \text{ } P_j :y.) \wedge \\ \neg \exists x, y (:x \text{ rdf:type } D_j. \wedge :y \text{ rdf:type } R_i. \\ \wedge :x \text{ } P_i :y. \wedge :x \text{ } P_j :y.)$$

そのマージ処理は、プロパティ P_i と P_j の両者をオントロジー O_3 において採用する。

2) 他方、プロパティ P_i が意味範囲においてクラス P_j と重なりがあるときには、下記のように表すことができ、

$$\exists x, y (:x \text{ rdf:type } D_i. \wedge :y \text{ rdf:type } R_i. \\ \wedge :x \text{ } P_i :y. \wedge :x \text{ } P_j :y.) \vee \\ \exists x, y (:x \text{ rdf:type } D_j. \wedge :y \text{ rdf:type } R_j. \\ \wedge :x \text{ } P_i :y. \wedge :x \text{ } P_j :y.)$$

マージ処理は、下記の選択肢1から3のいずれかとなる。

・選択肢1：マッピングの条件①から③を満たす、オントロジー O_1 のプロパティ P_i とオントロジー O_2 のプロパティ P_j の両者を採用する。なお、マッピングの条件④は満たす必要はない。

・選択肢2：プロパティ P_i を採用する

・選択肢3：プロパティ P_j を採用する

加えて、上記の選択肢1～3において、オントロジー O_3 におけるその時点のプロパティ P_k （オントロジー O_3 は順次更新される）について、意味範囲およびその定義域、値域において包含関係があるときには、下記のプロパティ間の関係づけを追加する。

$P_i \text{ rdfs:subPropertyOf } P_k$.

または、 $P_j \text{ rdfs:subPropertyOf } P_k$.

ただし、 P_k は上記の条件を満たす、意味範囲

において最下位レベルのプロパティとする。

4.4 プロパティのマージ：レベル2

調整（修正）は、プロパティの意味範囲、定義域、値域のすべて、またはそのいずれかの変更を指す。ラベル等の変更を含む。また、新規プロパティの追加とは、既存プロパティの調整結果に相当する新たなプロパティの追加を指す。あくまでも、マージ結果として先に採用したクラスに対応した調整とする。これは、プロパティの定義域、値域に指定されているクラス自体の採用が前提となるためである。

プロパティ P_i が意味範囲においてクラス P_j と重なりがあるときには、マージ処理は、下記の選択肢1から3のいずれかとなる。

・選択肢1：マッピングの条件①から③を満たすように、オントロジー O_1 のプロパティ P_i とオントロジー O_2 のプロパティ P_j の両者を等しく調整し採用する。なお、マッピングの条件④は満たす必要はない。

・選択肢2：プロパティ P_i を優先して採用し、それと整合するようプロパティ P_j を調整の上で採用する。

・選択肢3：プロパティ P_j を優先して採用し、それと整合するようプロパティ P_i を調整の上で採用する。

加えて、上記の選択肢1～3において、オントロジー O_3 におけるその時点のプロパティ P_k （オントロジー O_3 は順次更新される）について、意味範囲および定義域、値域において包含関係があるときには、レベル1の場合と同様、下記のプロパティ間の関係づけを追加する。

$P_i \text{ rdfs:subPropertyOf } P_k$.

または、 $P_j \text{ rdfs:subPropertyOf } P_k$.

ただし、 P_k は上記の条件を満たす、意味範囲において最下位レベルのプロパティとする。

引用文献

- 1) Shoichi Taniguchi. Mapping and Merging of IFLA Library Reference Model and BIBFRAME 2.0. *Cataloging & Classification Quarterly*. 2018, in press.
- 2) 谷口祥一. IFLA Library Reference Model と BIBFRAME 2.0 の統合：マッピングからマージへ. 2018 年度日本図書館情報学会春季研究集会発表論文集. 2018, p.57-60.
- 3) Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*, 2nd ed. Berlin, Springer, 2013, 511p.
- 4) Asunción Gómez-Pérez, et al. *Ontological Engineering*. Springer, 2004, 403p.
- 5) Heiko Angermann and Naeem Ramzan. *Taxonomy Matching Using Background Knowledge*. Springer, 2017, 103p.