

# 日本画用語事典の RDF 化によるオントロジー構築

小林美貴 (朝日新聞社)

mikobayashi@keio.jp

## 1. 研究背景と問題意識

近年、文化財の保存修復では科学的分析法の進展とウェブ技術の発展から、文化財の美術史的研究データ（制作者、画題、来歴等）と科学的分析データ（近赤外線画像、蛍光 X 線分析等）の異なる領域に分散して蓄積される情報の統合と分析、それに基づく保存修復処置が期待され、これを実現する仕組みとして、知識やデータの共通語彙を提供するオントロジーによる記述が望まれている。オントロジーは、知識に現れる概念と概念間の意味的關係を豊かな表現力で記述した概念体系であり、様々な領域で領域固有のオントロジーの構築が試みられている。文化財保存修復領域でも欧米を中心に構築例が見られるが、日本では実証的研究は少ない。

オントロジー構築法<sup>1)2)</sup>における「オントロジーの再利用」の観点から検討すると、文化財分野のオントロジーの大半が、文化遺産の概念モデルである CIDOC CRM<sup>3)</sup>（以下、CRM）を上位オントロジーに採用し、そのサブクラスやプロパティの規定化に取り組むという現状がある<sup>4)</sup>。一方、語彙の相互運用は、文化的相違に起因する多様性から実現が困難であるとの指摘もある<sup>5)</sup>。Odatらは、絵画保存修復情報オントロジーを CRM 等の拡張によって構築したが、用語集からの単語抽出に基づく概念の追加と領域専門家の評価及び調整を経て、既存のオントロジーに含まれない領域固有の概念構造を導き出した<sup>4)</sup>。

しかし、「情報資源からの情報抽出」の観点から、日本語の自然言語文を対象に概念及び概念間の關係を網羅的に抽出した先行研究は少なく、研究のほとんどが「語や語間關係」を論じた例である。このことから、日本画情報のオントロジー構築において、領域専門家が作成した情報資源を対象に、情報抽出に基づく概念体系の構築を行うことが有効であると考えられるが、構築方法とその実現可能性を検証する必要がある。

## 2. 研究目的と方法

### 2. 1 研究目的

本研究では、日本画の保存修復分野における日本画情報のオントロジーを、『図解日本画用語事典』<sup>6)</sup>のテキストからの情報抽出と記載事項の忠実な再現に基づいて試行的に構築し、汎用的な記述言語である RDF で記述し、さらに外部語彙とのリンク付けを行うことで意味的な相互運用性を検証する。特に、日本画情報のオントロジーを RDF クラスとプロパティという記述構文の側面から検討し、情報抽出に基づいたオントロジー構築の実現可能性と検討項目を明らかにする。

### 2. 2 研究方法

研究対象とする『図解日本画用語事典』は、東京藝術大学の保存修復日本画研究室が編集した、日本画に関する用語を収録した解説書である。用語は、伝統的な日本の絵画、日本画の流派、素材、道具、表現技法等の大きく 10 項目に分類される。保存修復専攻博士課程学生が修得する用語を網羅しており、オントロジー構築に適した情報資源であると考えられる。

本研究では、オントロジーの構築方法として、既存の情報資源である用語事典から情報抽出を行い、情報を忠実に構造化したモデル（概念体系）を導き出す方法を試行する。ここでは、Stuart のオントロジー構築方法<sup>2)</sup>のステップを踏まえながら、新たにフェーズ 1「トリプルの抽出」、同 2「RDF 化」、同 3「外部語彙とのリンク付け」の 3 つに再編した。フェーズ 1 では、事典のテキスト（図 1）を単語に分割し、語や属性、語間の關係を追加し、RDF 化を想定したトリプル（主語、述語、目的語）を抽出する。フェーズ 2 では、抽出したトリプルを RDF クラス、プロパティとして再度確認し、必要に応じて語の曖昧性解消や統制を目的としたクリーニングを行い、意味上の分類を行う。クラスやプロパティの定義（定義域と値域の指定を含む）を行い、最後に RDF グラフを作成する。フェーズ 3 では、クラスの一部を外部

語彙 (J-GLOBAL<sup>7</sup>, Getty Vocabularies<sup>8</sup>, CRM) と対応付け, ISO25964-2:2013 規格の「シソーラスのマッピング・タイプ」<sup>9</sup>のうち 3 種類の記号 (等価, 近似, 交差) を用いて照合し, オントロジーの拡張性を検証する。

|   |
|---|
| <p><b>緑青</b> ろくしょう   <i>rokushō</i>  </p> <p>緑色系顔料。岩緑青ともいう。原石は孔雀石 (マラカイト)。主成分は塩基性炭酸銅 [CuCO<sub>3</sub>·Cu(OH)<sub>2</sub>]。銅の鉱床の上部にできる。古くから東洋画の緑色顔料として珍重される。加熱すると深緑から黒緑、または黒色になる。</p> |
|---|

図1『図解日本画用語事典』テキストの例

### 3. オントロジー構築結果

#### 3.1 フェーズ1: トリプルの抽出

「主語」と「目的語」の抽出では, テキストを分割した単語が, 「述語」の抽出では, 作業者が追加した語や属性, 語間の関係が採用される傾向が見られた。

例えば, 見出し語「山水画」の解説文から抽出した結果では, テキストの分割語 96 語のうち 48 語 (50%) が採用され, そのうち 1 語が「主語」に, 44 語が「目的語」になった。一方, 「述語」には, テキストの分割語をそのまま採用する 경우가少なく (3 語), 作業者によって追加した属性・関係 41 語を採用した。

「主語」と「目的語」に採用された分割語には, 普通名詞 (例. 山水, 興隆) と時代や国, 人物名を表す固有名詞 (例. 六朝時代, 中国) が多く含まれた。一方, 「述語」には, 基本的に主語の属性や, 主語と目的語の関係を表す, 助詞をも含めた名詞や動詞等を抽出した。助詞を含む名詞・動詞等で表された述語は, (1) 動作・作用の主体や対象 (例. が描かれる, に描かれる), (2) 時空間上の始点や終点 (例. に変遷した), (3) 材料や仲介物 (例. を用いる) 等の関係を表現したものであった。

#### 3.2 フェーズ2: RDF化

RDF クラスとプロパティは, トリプルのクリーニングと意味上の分類を事前に行い, 定義した。特に, 「述語」の類似性による分類と統制は, (1) 異表記や同義の述語 (例. 「別称」と「異名」), (2) 意味上の類似性をもつ述語 (例. 「材料」と「材質」), (3) 階層関係にある述語 (例. 「支持体」と「基底材」) 等を対象に行なった。

ここでは RDF 化の結果から, 「伝統的な日本の絵画」と「日本画の流派」, 「素材」の項目別に, クラス階層と代表的なプロパティについて傾向を示す。

#### 3.3 「伝統的な日本の絵画」に属する用語

個別の用語 (見出し語) とその解説文が主たる参照対象となる。

- ・用語: 伝統的な日本の絵画の名称
- ・解説文: 名称の由来となった, 題材や絵画様式, 技法, 素材, 形態に関する情報

最終的に, クラス階層は最上位クラスを含めた最大 8 階層から構成され, 「伝統的な日本の絵画」に属する用語に対応する 55 クラスと解説文中から抽出した 139 クラスが配置された。ただし, そのうち 32 クラスが複数の上位クラスをもつという多重継承が示された。例えば, 「ngyc:絵巻物」は「ngyc:伝統的な日本の絵画」と, 「ngyc:表具の形式」の下位クラスである「ngyc:卷子」の 2 つのクラスを継承する (図 2)。こうした多重継承は, 本質属性をもつ基本概念と機能を表すロール概念を分離することで, 基本概念による属性継承に限った階層関係に変更するという考え方もあり<sup>10</sup>, 「ngyc:伝統的な日本の絵画」がロール概念で, 「ngyc:表具の形式」が基本概念と捉えることもできる。しかし, 本研究ではテキストからの情報抽出に即して配置し, 多重継承を許容した階層を導き出した。

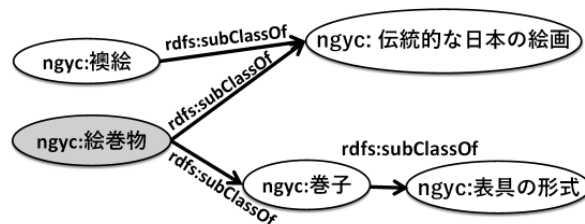


図2 「ngyc:絵巻物」の多重継承

プロパティには, (1) 単純な目的語インスタンスをとるもの (以下, 単純プロパティ) と, (2) 構造化した値をとるもの (以下, 構造化プロパティ) があり, 後者は空白ノードを用い, 複数の概念間の組み合わせを構造化して表現した。(1)には, 汎用的なプロパティである「rdfs:subClassOf」や「rdfs:label」が, (2)には, 独自語彙の「ngyp:制作」や「ngyp:変遷」等が該当する。

### 3.4 「日本画の流派」に属する用語

- ・用語：日本画の流派および絵画団体の名称
- ・解説文：絵師や画家、代表作の情報

この用語群に階層関係はほぼ認められず、最上位クラス「ngyc:日本画の流派」の直下に10用語に対応したクラスを配置する程度である。

単純プロパティでは、「rdf:type」や「ngyp:平仮名読み」、「rdfs:label」が頻出した。また、構造化プロパティである、「ngyp:継承者」と「ngyp:改称」は、RDF コンテナ (rdf:Seq) でグループ化し、コンテナのメンバーの「順序関係」(rdf:\_1, rdf:\_2, ...) を明示する方式で表した。例えば、用語「狩野派」の解説文には、後継者 (狩野元信, 狩野永徳, ...) の順序関係が記されており、RDF 化後もこれら順序を確認できるよう、本研究では、rdf:List に比べて簡潔なモデル表現となる、RDF コンテナを用いた。

### 3.5 「素材」のうち「顔料」に属する用語

- ・用語：日本画の色料 (絵具) の名称
- ・解説文：原料や材料とその成分の情報

クラス階層は、最大5階層からなる比較的簡素な階層構造であるが、14クラスとその上位クラス (系統で分類した色名の概念を含む) に多重継承が発生した。

単純プロパティは「rdfs:subClassOf」や「ngyp:平仮名読み」、構造化プロパティには、素材の製造にかかわる製造行為や材料、材料の状態や性質を組み合わせて表現する「ngyp:製造」を設けた。例えば、用語「緑青」の解説文を、「ngyp:製造」を用いて複数要素を関連性のあるものとして表現し、「ngyc:焼く」という製造行為と、それに伴う色料の「ngyc:変色」という状態変化、「ngyc:深緑\_顔料」から「ngyc:黒緑\_顔料」等への変色を示した (図3)。各要素は互いに異質の概念である。なお、「色」はある「実在物」の性質を表す概念と捉え<sup>10)</sup>、実在物 (例. ngyc:顔料) に依存する情報としてアンダーバーを用いて記述した (例. ngyc:黒緑\_顔料)。

ngyc:緑青 ngyp:製造 [a ngyc:製造; ngyp:製造行為 ngyc:焼く; ngyp:色料\_状態変化 ngyc:変色; ngyp:変色\_始点 ngyc:深緑\_顔料; ngyp:変色\_終点 ngyc:黒緑\_顔料; ngyp:変色\_終点 ngyc:黒\_顔料]

図3 「緑青」の解説文のRDFデータ (部分)

### 4 考察：RDFの記述構文にかかわる検討項目

オントロジー構築結果を踏まえて、次のRDFクラスとプロパティにかかわる検討項目と本研究における課題解決策を示す。

第一は、RDFクラスとプロパティの定義における記述の抽象度の調整である。本研究では用語事典の記載事項に忠実な再現を目指した結果、クラスにはテキストからの抽出語を採用し、プロパティには「ngyp:原料」や「ngyp:遺例」等の具体性を伴う独自語彙で、かつ同オントロジー内で適用可能な汎用性をもつ語彙を採用した。

例えば、解説文の“幾何学的な文様や人物、動物などが原初的な描写で描かれたものを装飾古墳といい”<sup>6)</sup>をRDFデータに変換する場合、クラスにはテキストの抽出語である「ngyc:装飾古墳」や「ngyc:人物」を採用した。一方、プロパティには、

(1) 抽出語の分類と統制を経て、定義した独自語彙の「ngyp:描く対象」等を採用、(2) より抽象度の高い語彙である「ngyp:制作行為\_対象」(プロパティ階層の上位プロパティとしての位置付け) を採用、(3) 外部の汎用的な語彙 (skos:related 等) を採用することが考えられる。比較検討した結果、テキストの記載事項により即したRDFトリプルによる文 (Statement, RDF文と略す) を作成できることから、(1) を採用した (図4)。

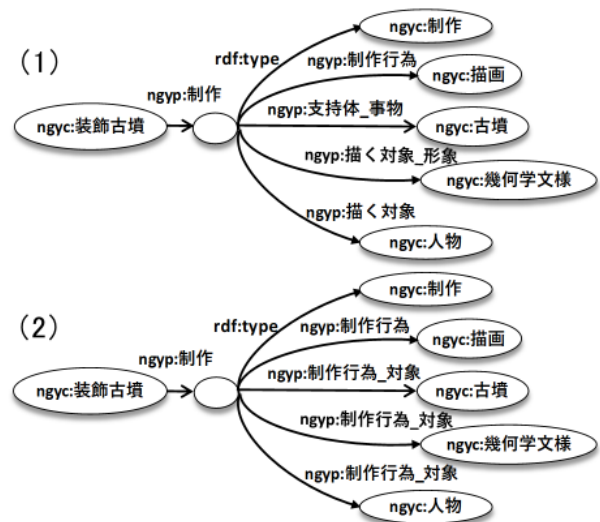


図4 「装飾古墳」の解説文に対するRDF化の2案 (部分)

さらに、RDF 文が等価の内容を示す場合、クラスには特殊性を、プロパティには汎用性を持たせるモデルを採用した。例えば、「和紙の原料は楮の靱皮繊維」を変換する場合、(1) 抽出語が示すクラス「ngyc:靱皮繊維」と特殊なプロパティ「ngyp:原料\_部分」を用いたモデル、(2) 特殊なクラス「ngyc:楮の靱皮繊維」と汎用的なプロパティ「ngyp:原料」を用いたモデルを検討し、(2) のモデルを採用した (図5)。

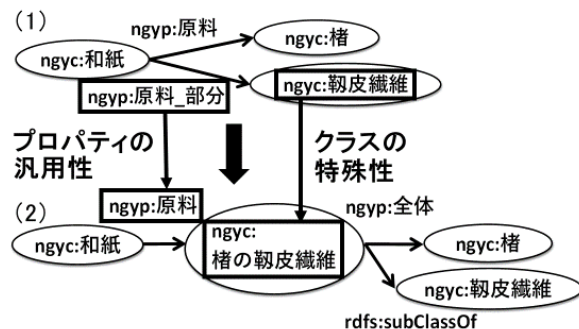


図5 クラスとプロパティの抽象度の設定

第二に、クラス階層における多重継承の許容である。重量オントロジーにおいては、1つの概念が複数の上位概念に対して `rdfs:subClassOf` で定義されるという多重継承は避けられ、概念が本質属性のみを継承するように設計されることが多い。しかし、特に学際的分野では、本質属性をもつ基本概念と機能を表すロール概念の混在がしばしば見られ、オントロジー構築において多重継承を許容することが多い<sup>10)</sup>。

第三に、構造化した値をとるプロパティの記述として、(1) 複数の互いに異質な要素を組み合わせて表現するもの、(2) RDF コンテナ (`rdf:Seq`) を用いたモデルで順序関係 (作業工程, メンバーの順序等) を表すものを採用した。これらプロパティは分散する概念同士を結びつけ、広がりのある概念ネットワークを形成する効果が期待できる。一方、プロパティを用いた複雑な構造と推論支援等の処理能力が相反関係にあることも留意しなければならない。

以上のように、用語事典からの情報抽出に基づきオントロジーを構築する場合、RDF の記述構文によるモデルは、記述の抽象度と自由度の高さによって概念間の関係を精緻に表現できる。特に日

本画保存修復分野のような、様式や主題、材料等の複数要素を含むオブジェクトを対象とする分野においては、多重継承をも許容する柔軟なモデル化が不可欠であるという結論に至った。今後の課題は、RDF 記述による表現の緻密性とデータ処理の容易性や推論等の実効性をバランスをもって整理し、再編成することである。

#### 参考・引用文献

- 1) Noy, N. F.; McGuinness, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001, p.1-25.
- 2) Stuart, D. "Chapter5 Building Ontologies". *Practical Ontologies for Information Professionals*. ALA Editions, 2014, p.97-136.
- 3) Le Boeuf, P.; Doerr, M.; Ore, C. E.; Stead, S. *Definition of the CIDOC Conceptual Reference Model*. ICOM/CIDOC Documentation Standards Group, CIDOC CRM Special Interest Group. Version 6.2.3. 2018-5. [http://www.cidoc-crm.org/sites/default/files/2018-05-16%23CIDOC%20CRM\\_v6.2.3\\_esIP%28XDP%29%28XM%29.pdf](http://www.cidoc-crm.org/sites/default/files/2018-05-16%23CIDOC%20CRM_v6.2.3_esIP%28XDP%29%28XM%29.pdf), (accessed 2019-10-20).
- 4) Odat, S. *A Semantic e-Science Platform for 20th Century Paint Conservation* (PhD Thesis). The University of Queensland, Australia, 2014. doi: <https://doi.org/10.14264/uql.2014.426>, (accessed 2019-10-20).
- 5) Chen, S.; Zeng, M. L.; Chen, H. *Alignment of conceptual structures in controlled vocabularies in the domain of Chinese art: a discussion of issues and patterns*. *International Journal on Digital Libraries*. 2016, vol.17, no. 1, p. 23-38.
- 6) 平山郁夫, 渡邊明義, 高田俊男, 田淵俊夫, 宮廻正明監修, 東京藝術大学大学院文化財保存学日本画研究室編. *図解日本画用語事典*. 東京美術, 2007, 221 p.
- 7) 国立研究開発法人科学技術振興機構. J-Global. <https://jglobal.jst.go.jp/>, (accessed 2019-10-20).
- 8) The J. Paul Getty Trust. *Getty Vocabularies*. <http://www.getty.edu/research/tools/vocabularies/>, (accessed 2019-10-20).
- 9) ISO 25964-2:2013: *Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies*. ISO, 2013, 99 p.
- 10) 溝口理一郎. *オントロジー工学の理論と実践*. 人工知能学会編. オーム社, 2012, 248p.