

貴重書中の部品を記述できるオントロジーに基づき Linked Data 化したメタデータを用いた人名ネットワーク構築の試み

Construction of a Human Network with Linked Data Based on a Bibliographical Ontology, Describing Components of Historical Rare Books

吉賀 夏子 *1 渡辺 健次 *2 只木 進一 *1
Natsuko YOSHIGA Kenzi WATANABE Shin-ichi TADAKI

*1 佐賀大学大学院工学研究科
Graduate School of Science and Engineering, Saga University

*2 広島大学大学院教育学研究科
Graduate School of Education, Hiroshima University

Digital collections of diverse cultural assets has been published on the Web. Current standard metadata sets for digital archives of cultural assets, however, have had various difficulties for researchers to retrieve targets efficiently. One of the reason of the difficulties is that bibliographical descriptions of cultural assets are not adequate to satisfy requirements of researchers in detail. We pick out historical rare books to construct a domain ontology based on a standard formal ontology. In this manner, historical rare books are described as both a book and a cultural asset. The ontology, especially, describes a historical rare book as a set of components expressed with terminologies of bookbinding. Each component has own author with its Unique Resource Identifier (URI). The URIs are references for unified database of authors. Retrieving the URIs enables us to construct a human network of authors in the target collection.

1. はじめに

Web 上では、多様な文化財資料に関する書誌メタデータ（目録）、画像および動画データが公開されている。これらの Web に蓄積された情報は、対象文化財群を比較分析する必要のある研究者によって共有され、新たな知識の形成に寄与する可能性がある。

しかし、実際の書誌から研究者が求める情報を得るまでに、研究者が払う負担は大きい。何故なら、あらかじめ博物館あるいは図書館等の専門機関で作成された書誌記述から、研究者が求める情報が存在するかを逐一読み取らなければならないためである。時には、研究者の勤と経験に頼らざるを得ないことも多い。

他方、研究者が各所で収集した書誌情報の断片を組み立て分析した結果が適切である場合、研究対象の文化財は再評価される。それは、一見私たちに何の関係もなく見える文化財が、私たちの生活や価値感にどのような影響を与えてきたのかを世の中の人々に理解させる、「価値あるもの」に変化することを意味する。

したがって、対象文化財に関する記述に、研究者が持つ知識とデータベース中の情報を容易に、矛盾なく組み合わせることが可能な枠組みがあれば、研究者の負担を軽減し、文化財の潜在価値をより多く引き出す可能性がある。

もちろん、博物館等の専門機関は、書誌記述に必要な要素について、対象物の専門家と検討した上で記述規則を既に制定している。しかし、従来の書誌メタデータを Web で検索利用する場合、各種の書誌記述がどのような意図で行われているかを、人に代わって機械的に認識することは極めて困難である。そのため、異なる文化財を通じた統一的な検索は困難である [安達 10]。

このような課題を解決するには、文化財の統一的な検索を行うために、あらゆる文化財で共通する書誌記述およびある文化財分野独自で用いられる書誌記述が、それぞれ機械的に明確になるような枠組みを書誌記述に導入する必要がある。さら

に、ある文化財で用いられる独自の記述を、研究者の関心に合わせて自由に設計可能な仕組みをつくる必要がある。そして、設計した書誌記述を Web で様々な立場の人々と共有することで、従来の専門組織による標準的な記述思想に研究者の記述思想が加わり、よりよい記述へ進化できる可能性がある。

以上のような柔軟な書誌記述を実現するために、これまで著者らは、主に江戸期以前の和古書、清代以前の漢籍が集合する貴重書コレクションの書誌記述に必要な概念モデルを実際に提案してきた。そして、文化財の記述に求められる枠組みについて模索してきた [吉賀 13, 吉賀 14, 吉賀 15]。貴重書は書籍であるが、その内容とともにモノとしての物理的な造りが注目される文化財としての一面を持つ [Lou 11]。そのため、多面的な書誌記述の構築方法を考察する場合に適している。

具体的には、貴重書のメタデータ記述に必要な機能要件を作成する。次に、国際組織が策定した標準の書籍および文化財用形式オントロジー [武田 04] を基に、貴重書分野に特化したドメインオントロジー（以降、提案オントロジーと呼ぶ。）の構築を通じてメタデータの設計図を作成する。

提案オントロジーは、貴重書を部品としての著作の集合体として捉える。貴重書の部品とは、表紙、見返し等を指す製本用語で指示されるものであり、書籍の部位を特定する。また、各部品の由来もオントロジーによる設計で明確にする。この部品単位の情報を、最重要事項を伝える際に必要な「5W1H」で取得可能にする。そして、同一タイトルの個別資料の比較を行う際、同じ部品同士での微妙な差異を比較可能にする。

最後に、実在するメタデータ（佐賀大学附属図書館市場直次郎コレクション [井上 07] 典籍（貴重書）の部）に対して、提案オントロジーに基づくメタデータを構成する。さらに、当該メタデータ（294 レコード、30987 トリプル）を、Linked Data 化する。

そして、属性および値で Web で一意の識別子である URI（Unified Resource Identifier）で定義可能なものには URI を紐付けるとともに、表形式であるメタデータをグラフ形式のメタデータへ変換する。

その結果、表紙、序などの部品を指定して検索すると、従来はテキストとしての部品名が入力されている箇所を結果として

連絡先: 吉賀夏子, 佐賀大学大学院工学研究科, 佐賀県佐賀市本庄町本庄 1, natsukoy@cc.saga-u.ac.jp

取得できるが、オントロジーを適用したメタデータでは、属性としての部品名を指定して検索可能となった。さらに、5W1H情報を含む部品を Web から URI で指定し、正規化した状態で取得可能となった。

特に、Linked Data 化により、メタデータ中に存在する人名に対して国立国会図書館の標目 URI[国立 11] を結びつけを行った。この人名に対する URI を利用することにより、コレクションに関わる人物のネットワークを、別名などによる名寄せを行いながら構築することができる。さらに、その人名ネットワークは、対象コレクションに限らずその周辺部まで拡張可能である。

本稿では、貴重書メタデータから人名ネットワークを構築する際に必要なメタデータ記述の基本的な枠組みについて、概要を説明する。

2. 提案オントロジー

2.1 概要

提案オントロジーは、書籍および文化財が融合した標準的な形式オントロジー FRBR_{OO}[Bekiari 13] を上位とする、貴重書分野に特化したドメインオントロジーである。FRBR_{OO} は、IFLA (国際図書館連盟) の書籍の内容を分析するための形式オントロジー FRBR_{ER}[和中 04] と ICOM (国際博物館会議) の文化財の成立過程を表現する CIDOC CRM[Crofts 11] が融合したものである。よって、FRBR_{OO} により、書籍も文化財としての記述対象として扱える。そのため、著者の創作内容のみでなく、編集、装丁、挿絵など書籍を成立に必要な要素も記述対象となる。

しかし、実際の書誌記述をオントロジーを使って表現するには、FRBR_{OO} を基に定義域、値域、使用回数、概念クラス間の関係を定めたドメインオントロジーを構築することが必須である。従来、FRBR_{OO} を用いて、手元に存在する文化財メタデータに適用する試みが行われてきた [Strle 12] が、ある分野の独自性と標準的記述を両立するための手段としてドメインオントロジーを構築するという手段は採られていなかった。

そこで、本研究では、貴重書を比較する研究者にとって必要な情報を取得するため、機能要件を定め、それを基に提案オントロジーをメタデータの設計図として構築した。

2.2 貴重書メタデータの機能要件

貴重書のメタデータ記述について、国立国会図書館および日本古典籍総合目録データベース [国文 06] の目録記述規則を調査した。その結果、専門家が対象文化財からメタデータを取得する際には、表紙、見返し、序等書籍中の様々な部位を単位として読み取ることが明らかとなった [増井 12]。

また、貴重書は現在流通している書籍の記述規則と貴重書のための細則で成立しており、言わば書籍として取り扱われている。

以上から、貴重書メタデータの機能要件を以下にまとめた。

1. 属性および値の正規化を行い、「部品」単位での著作情報を取得可能とする。
2. 表記ゆれが大きいメタデータに図書館の標目に対応する。
3. 貴重書を書籍に加え、文化財としても取扱い可能にする。

2.3 提案オントロジーの構成

提案オントロジーは、部品としての著作が書籍全体の情報に has-a 関係で繋がるように構築した (図 1 および図 2)。部品

とは表紙、序、本文等の書籍を構成する製本用語である。由来とは、各部位が人の手による創作表現か、版木による手刷り印刷による創作表現かを明確に分ける概念クラスの組である (図 2 の点線青枠内)。

また、部品のうち、必ず一つを冊子体とし、それには冊子体全体に関わる著作を表現した。

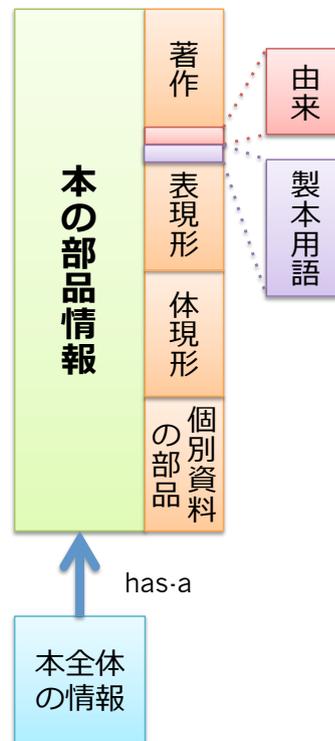


図 1: 図 2 の部品としての著作を表現するクラス群模式図

3. 提案オントロジーを適用したメタデータの有用性

提案オントロジーおよびそれを適用したグラフ形式のメタデータをメタデータ情報基盤サイト MetaBridge[メタ 13] に登録した。MetaBridge では、共用 SPARQL [DuCharme 11] エンドポイントが利用可能である。そこで、SPARQL エンドポイントでのクエリにより、2.2 節で示した機能要件を満たすか検証した。

その結果、製本用語を単位とする部品としての著作情報を URI を指定して取得可能となった。例えば、序文について記載された情報を検索する場合、従来の表形式メタデータでの検索 (表 1) では「序」あるいは「序文」という文字列がある該当レコードを列挙する。一方、グラフ形式のメタデータ検索 (表 2) では、属性の「序」のみを指定して、正規化した値のみを抽出可能となる。

さらに、貴重書では、同じ版木で印刷したものでも、印刷表現が異なるため、個別資料レベルで別個の扱いをする。そのため、各個別資料の同じ部品情報のみを URI で抽出し、詳細な特徴を画像と共に比較検討することは研究者にとって有用である。

表記ゆれに対しても、Linked Data により人名および分類名に標目 URI が紐付けられているため、その標目 URI を指

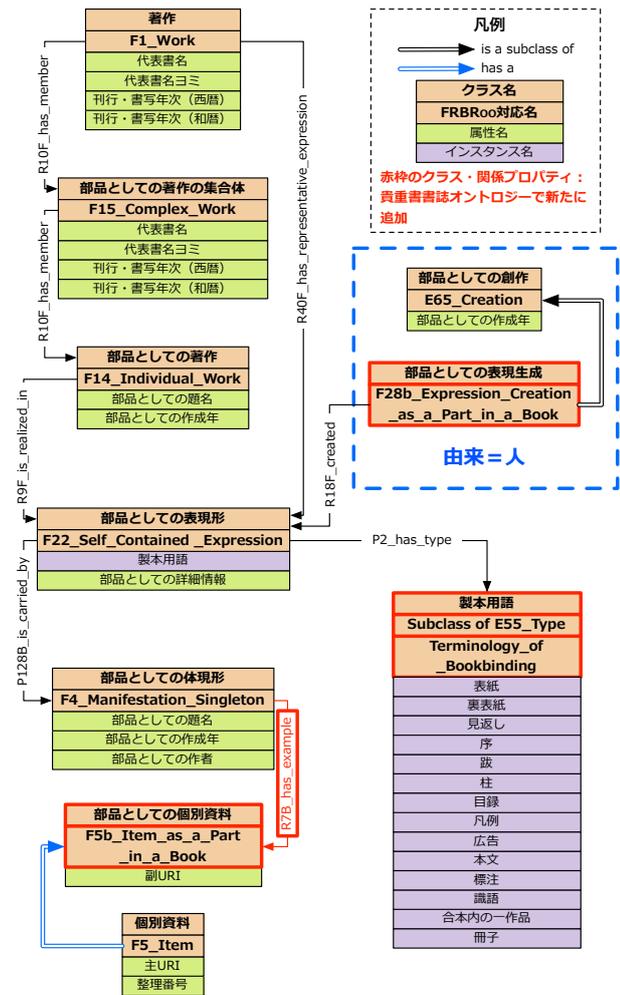


図 2: 部品としての著作を表現するクラス群 点線青枠内は部品としての著作を創作した由来(人)による表現生成過程を示す。

定して検索することで対応可能である。

このように、提案オントロジーを適用した貴重書メタデータは研究者が求める機能要件を満たしているため、有用である。

4. 提案オントロジーに基づく Linked Data を活用した人名ネットワーク構築

貴重書の研究において、その書籍成立の背景は重要なテーマである。書籍の来歴は、表紙、見返し、序、跋(後書き)等の部品の成立時期や関わった人物の情報から検討される。特に、序や跋の作者は、当該作品に大きな影響を与えた人物と考えられている。したがって、書物の各部品に関わった人物の関係をj知することは、書籍の来歴や他の書籍との関係を知る上で重要な情報である。

提案オントロジーに基づくメタデータでは、書籍の各部品に対して、5W1H 情報が対応付けられる。それらの情報には、URI を通じて NDLA (国立国会図書館典拠データ検索・提供サービス) [国立 11] の標目に対応付けられている。人名について考えると、NDLA の人名データは、貴重書に関わる人物の統一表記のデータハブである。すなわち、提案オントロジーに基づくメタデータ中の人物情報は、URI を通じて、人

表 1: 佐賀大学市場直次郎コレクション典籍(貴重書)の部メタデータにて序を全文検索した場合に抽出できる結果例(「序」に下線を付加)

書名	編著者名	刊行年	注記
西溪遺稿	高取九郎著 ・高取盛編	昭和 6 年	昭和 6 年紀元節後 1 日 74 翁草場謹庸 序 高取九郎・高取盛例言・ 佐賀晴雲堂本店写真印刷所。
殉難遺草	青雲閣城	明治 2 年	東山慨狂子 序。明治 2 年 1 月 青雲閣兼文 序。京都近江屋卯兵衛他 4 肆。
当世とらの巻(内)	田にし金魚	安永 7 年	安永 7 年一月田にし金魚 序
青楼心得辨	蓬萊山人	安政 4 年	見返題「女郎がひこゝるえ草」 安政 4 年一月蓬萊山人 序 *天明 5 年刊 「息子部屋」ノ改題本ナリ

表 2: 序の作成年および作成者抽出例

副 URI	作成年 1	作成者 1	作成年 2	作成者 2
?part	?date1	?name1	?date2	?name2
tenseki/182/8		"東山慨狂子"@ja	"明治 2 年 1 月"@ja	"青雲閣、兼文"@ja
tenseki/186/8	"昭和 6 年紀元節後 1 日"@ja	"翁草場謹庸"@ja		
tenseki/187/8	"安永 7 年 1 月"@ja	"田にし、金魚"@ja		
tenseki/189/8	"安政 4 年 1 月"@ja	"蓬萊山人"@ja		

(tenseki: <http://www.dl.saga-u.ac.jp/OgiNabesima/ichiba/tenseki>)

物データベースへと接続される。つまり、人名 URI を辿ることで、コレクション中の人物のネットワークを探ることができるだけでなく、NDLA の情報を使って、その周辺部まで探索が可能となる。

例えば、VIAF (国際バーチャル典拠ファイル) [Center 10] あるいは Wikipedia を経由して世界中の同義の標目 URI を検索し、部品としての著作情報の検索結果中に出現する序を記した人名(例えば、表 2 に出現する序の作成者)が、どのような社会の出来事あるいは著作に関わる人物と関係するかをさらに探索することが可能となる。つまり、提案オントロジーを適用することで、部品としての著作情報を抽出可能な Linked Data が文化財の貴重書と社会への関わりを解明する手掛かりになり得ることを示している。

5. むすび

本研究では、Web で多様な文化財の統一検索を可能にするため、書誌記述の在り方を見直した。そこで、書籍および文化財としての性格を持ち合わせる貴重書に焦点を当て、貴重書の書誌記述の枠組み作成を行った。

具体的には、貴重書を比較研究する研究者が貴重書検索に求める要素を機能要件として列挙し、それに基づき書誌記述の設計図となる貴重書分野に特化したドメインオントロジーを構築した。この提案オントロジーは、書籍および文化財標準である形式オントロジー FRBR₀₀ を基に、具体的な概念クラスおよび属性の使用を定めたものである。

次に提案オントロジーに実在する貴重書メタデータを投入し、提案オントロジーの有用性を検証した。その結果、貴重書研究者が求める要素である、部品としての著作情報を URI を指定し、必要な情報のみを正規化された状態で取得可能となった。

提案オントロジーをメタデータに適用する際に、Linked Data 化を行った。これにより、メタデータ中の値、特に人名につい

て、NDLA 標目 URI を同じく利用する Web 上のデータベースと対応付けが可能となった。この対応付けを探索することにより、部品としての著作に関わる人名に関連する付加情報を取得し、かつ様々なテーマによる人名ネットワークを形成して分析可能になる。コレクションに関わる人物のネットワークだけでなく、外部のデータベースを通じて、コレクション周辺の人物へと探索を拡大することも可能である。

コレクションに関わる人物のネットワークを知ることは、コレクションを構成する各書籍の来歴の一端を知ることである。つまり、各書籍の成立時期、影響を与えた人物の情報を通じて、書籍成立の背景を知ることができる。人物ネットワークから見える知見の整理、人物ネットワークの特性の分析などが課題となる。

参考文献

- [Bekiari 13] Bekiari, C., Doerr, M., Le Bœuf, P., and Riva, P.: FRBR object-oriented definition and mapping from FRBR_{ER}, FRAD and FRSAD (version 2.0), http://www.cidoc-crm.org/docs/frbr_oo//frbr_docs/FRBRoo.V2.0.draft_2013May.pdf (2013), 2015-01-31 参照
- [Center 10] Center, O. C. L.: *バーチャル国際典拠ファイル*, <http://viaf.org> (2010), 2014-10-10 参照
- [Crofts 11] Crofts, N., Doerr, M., Gill, T., Stead, S., and Stiff, M.: Definition of the CIDOC Conceptual Reference Model, http://www.cidoc-crm.org/docs/crm_version_5.0.4.pdf (2011), 2014-04-02 参照
- [DuCharme 11] DuCharme, B.: *Learning SPARQL*, O'Reilly Media (2011)
- [Lou 11] Burnard, L., O'Keeffe, K. O., Unsworth, J., 松原 良輔, 野中 進, 明星 聖子, 神崎 正英: *人文学と電子編集: デジタル・アーカイブの理論と実践*, 慶応義塾大学出版会 (2011)
- [Strle 12] Strle, G. and Marolt, M.: The EthnoMuse digital library: conceptual representation and annotation of ethnomusicological materials., *International Journal on Digital Libraries*, Vol. 12, No. 2/3, pp. 105–119 (2012)
- [メタ 13] メタデータ基盤協議会: *MetaBridge*, <http://metabridge.jp> (2013), 2014-04-02 参照
- [安達 10] 安達 文夫: *歴史研究情報の統合検索と歴史知識 (<特集> 歴史知識学)*, *人工知能学会誌*, Vol. 25, No. 1, pp. 17–23 (2010)
- [吉賀 13] 吉賀 夏子, 渡辺 健次, 只木 進一: *貴重書書誌の注記から抽出したメタデータによるオントロジー構築および書誌・美術関連 Linked Data と連携した検索システム構築*, *デジタル図書館*, No. 45, pp. 3–9 (2013)
- [吉賀 14] 吉賀 夏子, 渡辺 健次, 只木 進一: *書誌学的情報およびデータ入力を考慮した貴重書書誌オントロジーの構築*, 第 28 回人工知能学会全国大会 (JSAI2014), No. 1G5-OS-19b-1 (2014), 2014-12-20 参照
- [吉賀 15] 吉賀 夏子, 渡辺 健次, 只木 進一: *貴重書メタデータの設計図としての書誌オントロジーを適用した Linked Data*, 第 35 回セマンティックウェブとオントロジー (SWO) 研究会, No. SIG-SWO-035-08 (2015)
- [国文 06] 国文学研究資料館: *日本古典籍総合目録データベース*, <http://base1.nijl.ac.jp/~tkoten/about.html> (2006), 2014-04-02 参照
- [国立 11] 国立国会図書館: *Web NDL Authorities*, <http://id.ndl.go.jp/auth/ndla> (2011), 2014-04-02 参照
- [増井 12] 増井 ゆう子 (国文学研究資料館): *国文学研究資料館和古書目録データベースの作成*, <http://www.nijl.ac.jp/pages/event/seminar/images/wakosyodb.pdf> (2012), 2014-10-10 参照
- [武田 04] 武田 英明: *上位オントロジー*, *人工知能学会誌*, Vol. 19, No. 2, pp. 172–178 (2004)
- [和中 04] 和中 幹雄, 古川 肇, 永田 治樹 (訳): *書誌レコードの機能要件*, 日本図書館協会 (2004)