

# 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 Constructing a Bibliographical Ontology for Historical Rare Books, Concerning Actual Data Modeling and Editing.

吉賀 夏子\*<sup>1</sup>  
Natsuko YOSHIGA

渡辺 健次\*<sup>2</sup>  
Kenzi WATANABE

只木 進一\*<sup>3</sup>  
Shin-ichi TADAKI

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

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

\*<sup>3</sup> 佐賀大学大学院工学研究科  
Graduate School of Science and Engineering,  
Saga University

Digital archives for historical rare books have been rapidly growing and those bibliographical data have been also available through web services. It has been still difficult for novices in bibliography to search some bibliographical targets by using keywords. One of the reasons for those difficulties is that supplementary information, such as illustrations and author names of prefaces, which are important in bibliography, are not provided sufficiently in those search systems. In addition, librarians often needs to input bibliographical data from historical rare books by inferences, because parts of bibliographical data are not specified clearly. As a result, records contain marks of librarians' inferences, which prevent from users to acquire target results. To solve these problems, it is necessary to construct ontology for historical rare books describing the structure of books by selecting properties useful for users. Actual bibliographical data, "Ichiba Naojiro Collection", *ichiba*, owned by Saga University were used to extract new property-value relations. Some properties were separated into pairs of role and role-holder, for instance, a term for bookbinding and a title. After refactoring the properties by pairs of role and role-holder, ontology for historical rare books, based on FRBR<sub>OO</sub> (FRBR [Functional Requirements for Bibliographic Records]-object oriented), was constructed. This ontology composes a work as a book and some work(s) as a part of a book. Each part of a book clarifies a root of expression creation, such as a publication from a woodcut or an author. *Ichiba* were converted into RDF data by ordering the ontology rules of historical rare books. Some of those values, which contain the librarians' inferences are normalized with annotation properties, *skos:editorialNote*, remaining information of the inferences. The RDF data, terms and the ontology rules are registered on MetaBridge on the Web. RDF data stored in MetaBridge, driven by ontological rules, can be shared through a SPARQL endpoint. As a result, the search system, which works with the RDF data based on the constructed ontology, provides sufficient query functions by semantic analyses of the ontological rules.

## 1. はじめに

Web 上では古い貴重書を電子化した画像を閲覧できるデータベースが年々増加している。従来貴重書の研究者は実際に所蔵している施設に足を運び、所蔵品が破壊されないような厳重な制約の下で調査を行わなければならなかった。今日では、インターネット上に公開されている貴重書画像の閲覧を通じて、ある程度の事前調査を行うことが可能となっている。

一方、書誌情報の検索に目を向けると、書名や著者名などの限られたキーワード以外を用いた検索は、現代書籍を含めて、容易ではない。例えば、ある画家が挿絵を書いている書籍を、検索を通じて見つけることは容易ではない。このような情報は、通常の本誌情報データベースには含まれていないからである。

更に、貴重書の研究者が関心を持つ事柄は、現代書籍とは異なっている。貴重書においては、その書籍に関与した人が書籍中の「どこに」「誰が」「いつ」「何をしたか」や、版の変遷をたどることが重要となる[橋口 11]。ここで特徴的な事柄は、一つの書籍に関与する複数の人物、場所、行為の情報が必要とされていることである。例えば、保立は、書籍および書誌上に記載されているコンテンツのみを検索対象として捉えるのではなく、書籍が成立する過程を表現するオントロジー構築手法を確立して、歴史の本質を捉えるべきと論じている[保立 08]。しかし、今日の本誌情報データベースに対する検索では、挿絵や序文を書いた人名、版等に注目し、書籍を構成する部品に関する情報を、書誌学等に通じていない多くの利用者が容易に得て、書籍全体

の構造を把握することは難しい。

次に、貴重書データベースの提供者について考える。例えば、現代書誌ではほぼ必ず入力されている書名が貴重書の場合に虫食いや汚れ等で判読不能な場合が多い。更に、表紙に書名が書かれているとは限らず、見返し等別の場所に書かれた文を書名候補にする場合もある。また、入力にデータ編集者の推測(スキル)が必要な場面もあり、括弧書きやクエスチョンマークなどの文字列がデータ中に頻繁に見られる。検索可能なデータとするためには、このような編集者の考えを残しながらデータを正規化する必要がある。

上記のような困難さを解消するには、書誌利用者が必要とする貴重書用のメタデータを整理し、そのメタデータの構造および関連を示すオントロジーを構築することが一つの方法である。その上で書誌データ利用者のニーズに沿って貴重書に収められている内容を抽出し、データ利用者の認識可能な特徴を表す語彙を用いて、データを自由に抽出できれば、様々な研究への応用と発展が期待できる。

本研究では、貴重書のためのオントロジーを構築するため、まず佐賀大学貴重書コレクション[佐大 01]に含まれる「市場直次郎コレクション」[井上 07]の本誌データ(以降、*ichiba*)の注記の内容を分析した。これらの書籍成立過程において、序や跋(ばつ:後書き)の執筆者、外題と異なる内題および見返し題、出版書肆名等、貴重書利用者にとって重要な項目と値の関係を抽出した。抽出の結果、貴重書利用者は、書名、著者名、作成年といった一般的な項目の他に、版元に関する情報および誰が序跋(ばつ:後書き)文を書いたかというような、書籍の成立に関与する情報を求めていることが推測された。

連絡先: 吉賀夏子, 佐賀県佐賀市本庄町本庄 1 番地, 佐賀大学大学院工学研究科知能情報システム学専攻,  
12573015@edu.cc.saga-u.ac.jp

次に、元の書誌および抽出したメタデータを用いて、FRBR<sub>00</sub>[FRBR 13]を基に書籍の成立過程に主眼を置いたオントロジーの構築を行った。その際書籍全体、すなわち冊子としての著作を構成する最小単位の部品としての著作に分解した。冊子としての著作にまず主となる URI を割り当て、その主 URI に紐付く部品としての著作に対して従となる関係の副 URI を割り当てた。

さらに、データ編集者の推測を表す記号および文字類が含まれるデータには、編集者用のメモを意味するアノテーションプロパティをデータに付し、編集者の推測であるという情報を残しながら、検索に用いる Literal データをできるだけ正規化した。

加えて、メタデータ情報基盤サイトである MetaBridge [MB 13]に貴重書のためのオントロジーを具体的に適用した語彙定義、記述規則を登録した。これらの語彙定義および記述規則に沿って、*ichiba* メタデータから OpenRefine [OR 13]および RDF Refine[LiDRC 12]の RDF スケルトンを用いて作成し、RDF ストア(共用 SPARQL エンドポイント)に登録した。利用者は REST 方式による SPARQL クエリでこの RDF ストアに自由にアクセスし、書籍の構成要素、つまり部品としての著作情報を取得できる。

検索対象データを正規化された RDF 形式としたことで、検索システムは、本研究にて構築したオントロジーに従って検索を実行する。これにより、システム利用者は、書誌学者が行うように、冊子の構成要素にいたるまで、*ichiba* ドメインの全てのデータを検索対象とすることができる。また、そこに記述されている内容が、データ登録者の推測を含むものであるか否かを区別することができる。

## 2. 研究対象となる書誌データの概要

主として江戸時代に作成された書画、民俗資料等のコレクションである「市場直次郎コレクション」は、佐賀大学附属図書館所蔵の貴重書コレクションのひとつである。この内、典籍(書籍)においては 2014 年 3 月時点で 243 点の書誌データが順次公開され、全ページが電子化されている。その内 165 点が Web 上で閲覧可能となっている。

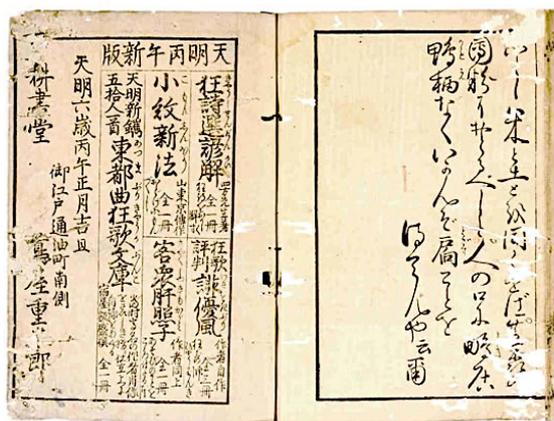


図 1 市場直次郎コレクションの典籍:巻末の目録および書肆(出版者)名

## 3. 新たな書誌項目の抽出およびデータの正規化

貴重書の書誌データである *ichiba* には多くの情報が含まれているにもかかわらず、データベースから求めるデータを取り出す処理が十分に施されていない。そのため、実在するデータを Linked Data[武田 13]に変換するために行った処理について以下に述べる。

### 3.1 新たな書誌項目の抽出

*ichiba* 書誌情報には、序文や跋文(後書き)が作成された時期、その作者名、挿絵の作者名、題名が抽出された製本用語上の場所、出版書肆名等多くの情報が、注記に補足情報として入力されている。すなわち、入力可能な書誌項目がないため、注記にダマダウンされていると言える。そこで、主に注記項目から新たな関係をあらかじめ手作業で抽出した[吉賀 13]。

その結果 *ichiba* の注記は、全レコードの 90%に含まれており、序跋文(前書きおよび後書き)の著者および作成年、出版書肆(出版者)名、合本中の題等、本文に関する補足というよりも、出版と書籍成立に携わる人に関する情報が多く含まれていた。新たに抽出した項目と値の関係は、貴重書のためのオントロジーを構成する属性として追加された。

### 3.2 貴重書におけるデータの正規化

貴重書の書誌データ入力は、くずし字等の判読はもとより、保存状態が良好とは限らない所蔵品を取り扱うため、客観的な情報を明確な文字列として入力するのが難しい。そのような場合は、データ入力者の推測や研究によってデータを補完する形で入力することがある。結果としてデータ中に推測を表す記号および文字列類が混じり、データが正規化されなくなる。したがって、キーワードを入力して全文検索したとしても、効果的に検索結果を得ることは難しい。

そのため、推測を表すデータには編集者のメモを表すアノテーションプロパティである *skos:editorialNote* を付加して、そこに推測であることを記入し、検索に用いるデータは指定されたデータフォーマットで入力することで正規化を行った。

## 4. 貴重書のためのオントロジー構築

### 4.1 FRBR<sub>00</sub> 概要および採用理由

1 章で述べたように、貴重書を通して、利用者が注目する歴史の分析を行うには、書籍の成立過程を表現できるオントロジーが必要である。貴重書の利用者は、書物としての性格だけでなく、外観や成立過程等、より文化財的な、モノとしての視点で貴重書を分析している。したがって、利用者が書籍のコンテンツにアプローチできるだけでなく、文化財としてアプローチできるようなオントロジーが必要である。

このような要求を満たすオントロジーとして本稿では、FRBR<sub>00</sub>に注目した。FRBR<sub>00</sub>は、国際博物館会議 ICOM ドキュメンテーション委員会 CIDOC が策定した、文化財用のオントロジーである概念参照モデル CIDOC CRM [CIDOC 11]と国際図書館連盟 IFLA の FRBR[和 04]が融合したオントロジーである。

FRBR<sub>00</sub>では書籍に含まれる著作を著作が成立するための最小構成単位である部品としての著作に分解できる。その部品としての著作の成立過程はある団体や人物による表現が実現されたものか、版元を印刷する際に、編集者によって生成された表現が実現されたものなのかを、その成立過程を含めて概念モデルとして表現できる。これらの理由から、FRBR<sub>00</sub>を本オントロジーモデルのベースに採用した。

### 4.2 FRBR<sub>00</sub> を拡張した貴重書のためのオントロジー概要

#### (1) 冊子体としての著作を部品としての著作に分割

実際は、FRBR<sub>00</sub>では概念クラスとその関連はクラス間の推移関係を含めたものが示されているが、そのクラスをどのように

組み合わせで、書誌世界のオントロジーモデルを構築するのは、利用者の手に委ねられている。

そこで、図2、図3および図4で示すような、FRBR<sub>00</sub>のクラスを一部拡張した貴重書のためのオントロジーを構築した。具体的には、ある書籍(冊子体)としての著作に対し、まず主URIを付与した。その冊子体内では、書籍の部位を表す製本用語を用いて部品としての著作に分け、各著作に副URIを付与した。

この時、FRBR<sub>00</sub>では表現形および表現形で人によるものと印刷によるもののクラス群を分離しているが、表現生成および個別資料のクラスには特に印刷用のクラスを指定していない。そのため、新たに印刷用と人用クラスに分離した。

これにより、部品としての著作を成立させ、その表現を生成したエージェントは何か、すなわち著者としての人物によるものか、版本を用いた印刷によるものかを明確にすることができた。

(2) 概念クラスへの各属性の割り振り

本オントロジーを構成する各クラスに属する項目はFRBR<sub>00</sub>のクラス説明に沿って割り振った。

まず、F5\_Item(個別資料)には主URIが割り振られる(図4)。F5はE24\_Physical\_Manmade\_ThingのサブクラスであるRare\_Bookのサブクラスである。Rare\_Bookには書物の分析に必要な属性および管理情報が含まれている(図2)。

| Rare_Book: Subclass of E24_Physical_Manmade_Thing |      |       |            |
|---|------|-------|------------|
| 冊子体としての主URI                                       | 刊号   | 画像URI | コレクション名    |
| 冊子体としての整理番号                                       | 注記   | 媒体    | 文庫         |
| 代表書名  | 印刷時期 | 原著者   | 他          |
| 代表書名ヨミ  | 印刷手法 | 原本発行日 | 所蔵者        |
| 冊子体としての編者   | 装丁   | 改題名   | 法人名        |
| 刊行・書写年次(西暦)                                       | 蔵版   | 改題年   | 法人名URL     |
| 刊行・書写年次(和暦)                                       | 書肆   | 関連本   | 図書館名       |
| 編集情報  | 編    | 広告    | 部品としての副URI |
|   | 補    | 書型    | 製本用語       |
|   | 書写   | 巻冊    | 部品としての題    |
|   | 校閲   | 印記    | 部品としての作成年  |
| 管理情報  | 翻訳   | 言語    | 部品としての詳細情報 |
|   | 板行   | 分類    | 部品としての詳細情報 |
|   | 後印   | 画像公開  | 画像位置       |

編集者情報: 「編集上の役割名称」 太字属性: ichiba基本属性 □属性: 冊子に関する属性  
 クラスのインスタンスおよび人名情 標準書体属性: ichiba拡張属性 □属性: 部品に関する属性  
 報で構成 □属性: 各属性に紐づくアノテーション □属性: 管理情報および来歴

図2 E24\_Physical\_Manmade\_ThingのサブクラスRare\_Bookに含まれる属性

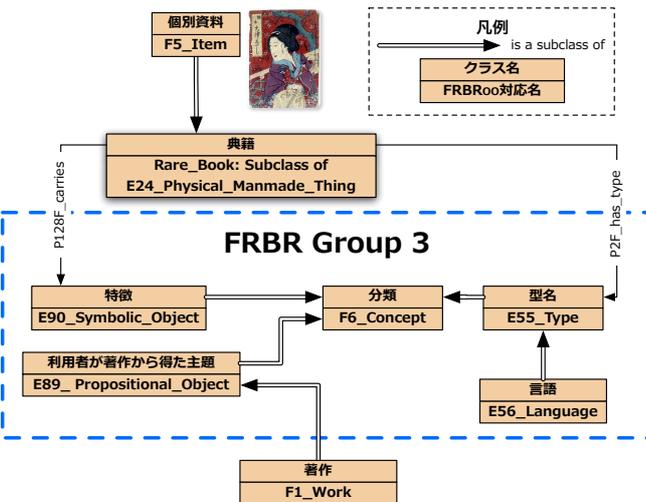


図3 E24のサブクラスRare\_Book, FRBR Group3およびF1\_Workの関連図

そして、E24 から E90\_Symbolic\_Object(書物の特徴)、E55\_Type(ドメインや制約を表現する型名クラス)が導き出され

る。また、図3の通り、E89\_Propositional\_Object(利用者が著作から得た主題)はある著作の主題を構成する F1\_Work のスーパークラスである。書物の分類である F6\_Concept は E90, E55, E89 の集合である。

E89 のサブクラスである F1\_Work, F15\_Complex\_Work(冊子体としての著作)および F14\_Individual\_Work(部品としての著作)群は、題名とその成立年の属性でシンボル付けされる(図4)。

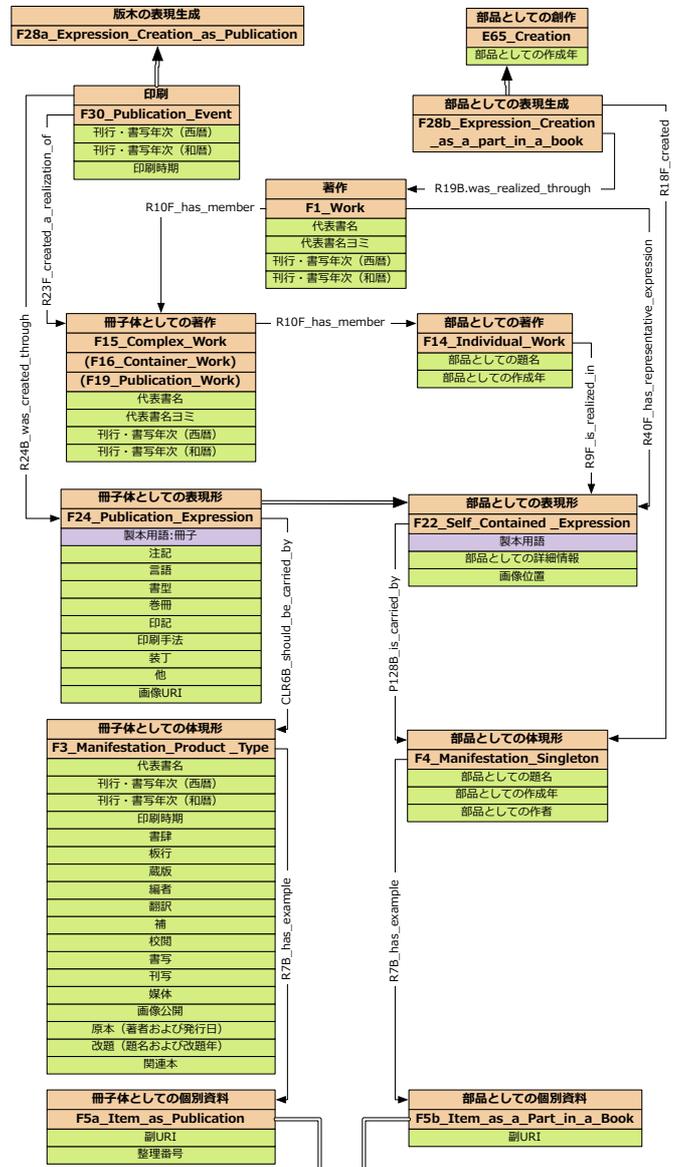


図4 冊子体としての著作由来の F5a および部品としての著作由来の F5b に副URIが付けられて F5\_Item(最下部)に集約する概念モデル図

これらの著作クラス群を実現する表現形である F24\_Publication\_Expression および F22\_Self\_Contained\_Expression には、可視的情報、言語、注記といった客観的情報を司る属性群およびその表現位置である製本情報のクラスから生成されたインスタンスが設定される。

これらの表現形の生成イベントクラス群 F30\_Publication\_Event あるいは E65\_Creation を発動する元と

なるクラスは、版木を用いた印刷または手書きによる書写、あるいは部品としての表現形を生成した人となる。

表現形である `F3_Manifestation_Product_Type` および `F4_Manifestation_Singleton` には、表現形が実際に収められている容器を示すため、書名と成立(作成)年および媒体、編集、出版に関する属性が用いられる。

`F5a_Item_as_Publication`, `F5b_Item_as_a_Part_in_a_Book` はそれぞれ表現形を例示する、冊子体としての個別資料および部品としての個別資料である。`F5a` および `F5b` は主 URI を持つ `F5_Item` のサブクラスであり、各々に副 URI を割り当てるため、本オントロジーに追加した。

### (3) ロール概念を用いた項目のリファクタリング

項目そのものについても、表紙の題、見返しの題、章題といった属性は、ロール概念+ロールホルダーの考え方で項目を分解できる[溝口 06]。具体的には、製本用語+題・著者・詳細、または編集上の役割+人名というように項目を分解した。また、`E55_Type` のサブクラスとして、`Terminology_of_Bookbinding` (製本用語)および `Editorial_Role` (編集上の役割名称)を新たに設定し、それらのクラスのインスタンスとして設定された用語や役割名を属性の値として用いた。

### 4.3 MetaBridge を利用した語彙定義記述規則および RDF ストアへの生データ登録と利用

貴重書のためのオントロジーに対応した *ichiba* の作成は以下の方法で行った。

MetaBridge では、各方面の語彙定義および記述規則を比較的容易に登録し、RDF に変換された生データを RDF ストアと呼ばれる共用 SPARQL サーバに蓄積して公開できる。

語彙定義については、Protégé[BMIR 14]でクラスとクラス間の関連を示すオブジェクトプロパティを設定した後、属性名とその定義域および値域をデータプロパティに入力し作成した。記述規則は、MetaBridge が提供する簡易 DSP という、OWL 記述をより平易に入力可能なテンプレートを用いて作成した[MB 11]。

さらに、作成した語彙定義と記述規則に従い、OpenRefine の RDF スケルトンを用いたメタデータマッピングと生データの RDF 化を行った。RDF スケルトンとは、OpenRefine に RDF Refine 機能拡張を付加することで利用可能なメタデータマッピングツールである。その後、生成された RDF データを RDF ストアに登録し、外部から REST 方式で SPARQL 問い合わせ可能にした。

### 5. まとめ

書籍の成立過程を明らかにし、その情報を抽出可能な、貴重書のためのオントロジーを構築した。すなわち、ある貴重書の著作を冊子体としての著作および部品としての著作に分解し、部品としての著作の表現生成および表現形について、印刷あるいは人由来かを明確にした。冊子体としての著作には主 URI、それに紐づく部品としての著作に副 URI を割り当てた。部品としての著作に URI を設定することで、例えばある人物が作成した挿絵の情報を SPARQL サーバのクエリを用いて著作単位で引き出すことが可能となる。

構築したオントロジーを構成するクラスには、対応する属性を割り当てた。貴重書の書誌項目に対し、直接 Dublin Core などの汎用メタデータスキーマおよびアプリケーション・プロファイルを利用すると、貴重書の成立過程を詳細に分析する必要がある利用者にとっての概念表現が不足する。しかし、冊子体としての著作を部品としての著作に分解し、さらに、項目を意味の上で分解し構造化すると、汎用メタデータの適応が容易となった。

実際に、*ichiba* の生データを本オントロジーに対応させるため、MetaBridge 上に公開した。まず、編集者用のメモを書き込むアノテーションプロパティを用いて、推測を表す記号などを値から除去するなどして、生データをできるだけ正規化した。次に、MetaBridge に語彙定義および記述規則を作成し、登録した。さらに登録した記述規則をもとに生データを RDF に変換した。この RDF データを MetaBridge の RDF ストアに蓄積し、REST による SPARQL クエリの結果呼出に対応させた。

すなわち、本オントロジーの記述規則に沿った検索結果を返すことが可能なデータを作成した。

今後は貴重書のためのオントロジーを元 RDF に変換されたデータに対応する、検索システムの作成および本オントロジーを用いたデータの有効性を検証する予定である。

### 参考文献

- [橋口 11] 橋口侯之介: 江戸の本屋と本づくり 続和本入門, 平凡社, 2011.
- [保立 08] 保立道久: 歴史のひろば: 歴史データベースの将来と歴史知識学: コンピュータはただの便利な道具か, 歴史評論, 702, pp. 80-89, 2008.
- [佐大 01] 佐賀大学附属図書館: 佐賀大学貴重書コレクション, <http://www.cc.saga-u.ac.jp/OgiNabesima/>, 2001. (2014-03-03 参照)
- [井上 07] 井上敏幸(編): 市場直次郎コレクション目録, 佐賀大学附属図書館・地域学歴史文化研究センター, 2007.
- [FRBR 13] International Working Group on FRBR and CIDOC CRM Harmonisation: FRBR object-oriented definition and mapping from FRBR<sub>ER</sub>, FRAD and FRASD (version 2.0), [http://www.cidoc-crm.org/docs/frbr\\_oo/frbr\\_docs/FRBRoo\\_V2.0\\_draft\\_2013\\_May.pdf](http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V2.0_draft_2013_May.pdf), 2013. (2014-03-03 参照)
- [MB 13] 一般社団法人メタデータ基盤協議会: MetaBridge, <http://www.metabridge.org/>, 2013. (2014-03-03 参照)
- [OR 13] OpenRefine Community: OpenRefine, <http://openrefine.org/>, 2013. (2014-03-03 参照)
- [LiDRC 12] Linked Data Research Centre (LiDRC): RDF Refine, <http://refine.deri.ie/>, 2012. (2014-03-03 参照)
- [武田 13] 武田英明, 大向一輝, 加藤文彦, 他(訳): Linked Data Web をグローバルなデータ空間にする仕組み, 近代科学社, 2013.
- [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.
- [CIDOC 11] Definition of the CIDOC Conceptual Reference Model, ICOM/CIDOC CRM Special Interest Group, [http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.0.4.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf), 2011. (2014-03-03 参照)
- [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.
- [溝口 06] 溝口理一郎(編): オントロジー構築入門, オーム社, 2006.
- [BMIR 14] Stanford Center for Biomedical Informatics Research: Protégé, <http://protege.stanford.edu/>, 2014. (2014-03-03 参照)
- [MB 11] 一般社団法人メタデータ基盤協議会: メタデータ情報共有のためのガイドライン, <http://www.mi3.or.jp/item/A03.pdf>, 2011. (2014-03-03 参照)