

平成 30 年 5 月 22 日現在

機関番号：32612

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330369

研究課題名(和文) RDA実装に向けたモデリングの適用およびメタデータ作成支援システムの開発

研究課題名(英文) Conceptual modeling of RDA for its implementation in Japan and system development to assist in metadata creation with RDA

研究代表者

谷口 祥一 (TANIGUCHI, Shoichi)

慶應義塾大学・文学部(三田)・教授

研究者番号：50207180

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：新たなメタデータ作成ガイドラインであるRDA (Resource Description and Access) のわが国での実装に向けて、メタデータ作成者による理解を支援するために、RDAに即したモデルとRDAメタデータ作成タスク自体のモデルなどを構築して提示し、加えて、RDAメタデータ作成スキーマとして有力視されているBIBFRAMEの検討を行った。また、RDAに基づくメタデータ作成処理の支援を意図して、既存書誌・典拠レコードから著作と表現形のデータを抽出し統合すること、およびメタデータの主題表現力の強化に向けた分類と件名の統合的活用のための状況調査と実験等を行った。

研究成果の概要(英文)：For implementing RDA (Resource Description and Access), which is a new guideline on metadata creation in libraries, conceptual modeling for the bibliographic universe represented in RDA instructions and also modeling for the description tasks making up RDA metadata creation were conducted. BIBFRAME, which is a primary candidate schema for RDA metadata, was evaluated from the viewpoint of suitability for RDA metadata. In addition, for aiming to assist in metadata creation with RDA, extracting work and expression data from existing bibliographic and authority records integration was experimented. Integrated utilization of classification schemes and subject headings was also experimented towards the enforcement of subject representation in RDA metadata.

研究分野：図書館情報学

キーワード：メタデータ RDA FRBR 概念モデル BIBFRAME

1. 研究開始当初の背景

(1)2010年にRDA(Resource Description and Access; 資源の記述とアクセス)が冊子体およびオンラインツールRDA Toolkitとして刊行された。RDAはこれまでのAACR2(英米目録規則第2版)の後継に当たるが、欧米における長期間の議論と検討の末に完成を見た、図書館等における新たなメタデータ作成ガイドラインである。米国議会図書館や書誌ユーティリティOCLCなどは、2013年春より正式にRDAを採用しており、他の主要国でも同様に採用が計画されていた。他方、わが国では、国立国会図書館が国外刊行物(いわゆる洋資料)に対してRDAの適用を開始した。

(2)RDAは、これまでのAACR2とは異なるいくつかの大きな特徴を持っており、FRBR(書誌レコードの機能要件)、FRAD(典拠データの機能要件)という概念モデルに完全に立脚している。著作、表現形、個人・家族・団体という、これまでの典拠データに相当する部分について明確な位置づけと規定化を図っている。著作間、表現形間、体現形間の関連や、著作等と個人等との関連など、関連の同定とその記録を重視するなど、複数の特徴を備えている。このような新たな特徴群を有するため、RDAを正確にかつ短時間で理解することは容易とは言い難い。目録担当者向けの学習用教材なども公開されているが、それらをもって十分とは考えにくい。また、実際にRDAを用いてメタデータ作成を行う際には、例えばこれまでと大きく異なる処理内容を伴う、著作や表現形を記述し関連づける処理において戸惑うことが予想された。

(3)こうした点から、メタデータ作成者の理解支援に向けて、RDAに即したモデルおよび適用事例等を効果的に提示すること、およびRDAに基づくメタデータ作成処理の支援システムを開発することが必要かつ有効と認識された。その背景には、研究代表者自身が、RDAの刊行以前からFRBR等の概念モデルを独自の視点から検討してきており、RDAの完成後は主にFRBRおよびFRADとRDAの概念モデルとしての相違を明らかにすることをやってきたことがある。

(4)特にわが国の既存の書誌・典拠レコードは、日本目録規則の伝統や慣習もあり、欧米に比べて著作・表現形の記録が脆弱である。それゆえ、RDAによるメタデータ作成を円滑に進めるためには、新規の処理対象資料に対する著作・表現形の同定と既存の記録化された著作・表現形との照合処理を適切に支援する、独自の機能が必要であった。

2. 研究の目的

(1)RDAに即した概念モデルの提示：RDAというガイドライン自体が表す概念モデルを、FRBRおよびFRADとの相違を明確にしつつ明らかにする。また、FRBRなどには含まれて

いない要素、例えば「参照した情報源」「目録担当者注記」など、いわばメタレベルにある要素を含めて妥当なモデルを提示する。こうした検討には既に着手しており、全体の整合性などをさらに整え、包括的なモデルとして提示することを試みる。

(2)RDA ルールの分析とモデリング：メタデータ作成者にかかわるタスクのモデリングを行い、このタスクとRDAが示す記述原則(Principles)とから、RDAガイドラインを構成する個別ルールの位置づけを再規定する。他方、個別ルールの構成要素等を分析し、その内的な構造化を図るとともに、ルール間の関係分析を行い、個別ルール内部およびルール間に関わる構造の可視化を図る。これらは、RDAによる記述処理の理解を容易にすることに通じると考える。

(3)ルールの適用事例の蓄積・表示機能をもつシステムの構築：メタデータ作成処理において入力データとなる情報源および適用した個々のルール群、そして処理結果であるメタデータ(記録されたエレメントとその値)を組み合わせて蓄積する機能を有するシステム、さらには蓄積されたメタデータとその情報源、適用ルールを適切に表示する機能を有するシステムを構築する。

(4)既存レコードからの著作・表現形データの抽出・統合システムの開発と評価：RDAに基づくメタデータ作成処理のうち、特に著作と表現形のメタデータ作成を支援するために、既存の書誌・典拠レコードから著作および表現形に該当するデータを機械的に抽出し、複数レコードから抽出されたそれら同一著作・表現形のデータの統合化を図るシステムを開発し、その性能を評価する。

3. 研究の方法

(1)RDA自体が表す概念モデルを、FRBRおよびFRADを参照しつつ、それらとの相違を明確にした上で構築した。加えて、メタレベルにあるRDA特有の要素をも整合的に表現できる方式を模索した。次に、こうして構築したモデルを踏まえて、さらに構造的な値の表現力の導入など、詳細化・精緻化を図る方策を検討した。その過程では、他の複数のメタデータモデル等を参照した。

(2)RDAモデル構築の一環として、RDAおよびそれが依拠するFRBRを、優先する実体の観点から改めて検討した。研究代表者は以前に「テキストレベル実体を基盤にした概念モデル(表現形優先モデル)を提案している。そこでRDAやFRBR等に対して、優先する実体を変化させたときのその帰結を確認しつつ、特に表現形を優先したときのモデルを検討した。

(3)個別のRDAルール(インストラクション)

を分析しモデリングを図る一つの方策として、メタデータ作成者にかかわるタスクの観点から RDA ルールを分析し、記述処理タスク自体のモデリングを試みた。また、設定した記述処理タスクに基づき、個別ルールの RDA 内での位置づけの明確化と、他のルールやタスクとの相互関係の明確化を図った。

(4)米国議会図書館が主導する BIBFRAME プロジェクトにおいて提案された最新版のバージョン 2.0 (2016 年公表)を取り上げ、そのモデルおよび語彙について検討した。BIBFRAME は、現在、RDA を用いたメタデータ作成に際して利用しうる有力なスキーマとして提案されているため、RDA モデルとの相違の確認、RDA メタデータ作成用スキーマとしての適切性などを検討した。

(5)国立国会図書館件名標目表 (NDLSH) において、件名標目に対して概念上で対応する日本十進分類法 (NDC) 新訂 9 版の分類記号が「代表分類記号」として示されている。国立国会図書館作成の書誌レコードに付与された NDLSH 件名標目と NDC 分類記号の組み合わせの中で、代表分類記号が付与されている割合を調査した。普通件名と地名による件名に分けた集計、分類記号の前方一致を加えた集計、レコード内先頭出現の件名標目に限定した集計など、詳細な集計と分析を行った。また、「代表分類記号」が付与されている件名標目の範囲は限られており、細目を伴った件名 (主標目 + 細目) には代表分類記号が原則付与されていない。そこで、書誌レコードに付与された細目付き件名と NDC 新訂 9 版の分類記号の組み合わせの中から、当該件名の代表分類記号となりうるものを機械学習によって同定することを試みた。人手によって妥当性を判定した件名・分類記号ペアを準備し、機械学習の適用実験を行った。なお、こうした調査および実験は当初の研究目的の範囲に設定していなかったが、RDA メタデータの主題表現力を最終的に強化するためには不可欠と考え、敢えて追加した。

(6)RDA に基づく著作と表現形のメタデータ作成を支援するために、国立国会図書館作成の既存書誌・典拠レコードから著作および表現形に該当するデータを機械的に抽出し、複数レコードから抽出されたそれら同一著作・表現形のデータの統合化を図った。現行の MARC21 書誌・典拠レコードスキーマを採用するのではなく、著作・表現形・体現形・個別資料ごとのメタデータスキーマによるメタデータ作成を前提とし、どの程度の網羅性で著作・表現形のデータ (エレメントとその値の組) を抽出できるのか検証を試みた。

4. 研究成果

(1)RDA は FRBR および FRAD に依拠しつつも、独自の概念モデルを有していると捉えるべきことを検証し、その成果を RDA モデルとして公

表した。加えて、さらなる詳細化・精緻化を目的として、イベント実体の導入と、表示要素に対応する属性の独立実体化を図り、それらを組み合わせ採用した RDA モデルを提案した。

「イベント」とは、イベントへの入力と出力となるリソース (実体) 関与するエージェントとなる実体とそれぞれ関連をもち、時間や場所などの独自の属性を備えたものとなる。こうしたイベント実体の導入により、リソースとエージェント等の実体間の、全体としての多項関連を表現できる。タイトル表示など、情報源上の表示要素でもある一部の属性 (RDA エレメント) について、独立した実体とすることを提案した。それにより、たとえば体現形とそのタイトル、責任表示、出版地、出版者名、出版日付等の値 (表示文字列) との多対多の関連、および情報源上での各種表示と最終的に選択され記録された値との関係表現することができる。併せて、メタデータとそのモデルの構造化、特に構造的表現力の観点に依拠し、RDA モデル、BIBFRAME モデル、FRBRoo (オブジェクト指向 FRBR) などを取り上げ、それぞれの構造化の方式や程度について検討した。これらに該当する一連の研究成果は、大きな意義をもち、かつ有効と判断される。

(2)同一実体構成によるモデル内で特定実体を優先したときに、どのような変化が生じるのか、さらにはいずれの実体を優先すべきかは、相互運用性にとって重要な問題であることを再確認した。RDA およびそれが依拠する FRBR モデルの実体構成を素材にして、その優先する実体異なる 5 つのモデルパターンを派生させ、モデル間に生じる相違を確認した。各実体のインスタンスの必須度や、個別資料に出現したタイトル等を属性として割り当てる実体などの相違が、特定の実体を優先することの直接的な帰結であり、個人等の実体との関連づけ、集合的実体に相当するケースの扱い、縮約・改訂等の扱い、同一コンテンツの場合の扱いなどに生じる相違が間接的な帰結であると結論づけた。

(3)RDA メタデータ作成者によるメタデータ作成 (記述処理) という、これまで取り上げられなかったことのないタスクについてモデル化を試み、記述処理タスクのモデルを提案した。個々のタスクは「イベント」概念に相当すると捉えることができ、多様な要素を結びつける多項関連をイベント実体をもって表現した。タスクのモデルは基本的にタスクへの入力と出力のリソース、処理の実行者である「カタログガー」そして処理を制御する「RDA インストラクション」などから構成される。FRAD モデルから典拠データ作成タスクのモデルを導出し、RDA に適応するよう修正を図った。また、RDA 記述処理タスクは複数の粒度で設定し、たとえば体現形の同定・記録タスク、体現形タイトルの記録タスク、体現形タイトルに関わる個別インストラクションに対応するタスクなどのモデル化を行った。なお、こうしたモデリング結果を活かして RDA

ルールの適用事例の蓄積と表示機能をもつシステムを構築する予定であったが、主に時間的な制約により未着手となった。

(4)BIBFRAME の最新バージョン 2.0 について、主に RDA メタデータの作成用スキーマとしての観点、および RDA メタデータを超えて多様なメタデータの交換・共有用スキーマとしての観点の両方から、その適切性を検討し、問題点等を指摘した。 については、現行の BIBFRAME では RDA エレメントに対応するプロパティが十分ではなく、RDA から多対一または多対多の対応となるプロパティが相当数あり、これらは RDA のエレメントとその値の組みに適切に戻すことが困難であること、また表現形が適切に扱えないなど、RDA メタデータを十分に表現できないことを指摘した。 については、プロパティの定義域が極めて緩やかな設定とされている結果、一定程度の品質を伴ったメタデータの管理に適切ではないことなどを主に指摘し検証した。これらの成果に加えて、RDA や BIBFRAME を含め、図書館目録全体のこれまでの経緯および現在の動向等について、図書の実担執筆および雑誌論文の執筆、学会の招待講演等を機会にして広範にまとめることを行った。これらにより、RDA そして BIBFRAME の位置づけがさらに明確にできたものと考えられる。

(5)RDA メタデータの主題表現力を最終的に強化することを意図して、NDLSH 件名標目と NDC 分類記号の対応づけの状況調査、および両者の統合的活用法の検討を行った。普通件名と地名による件名に分けた集計、分類記号の前方一致を加えた集計、レコード内先頭出現の件名標目に限定した集計など、さらには NDC 分類記号の側からの集計など、詳細な集計と分析を行った。また、両者の対応づけの自動化または作業支援の方策として、細目付き件名標目に対応する分類記号を機械学習によって自動的に付与できるのかを実験により検証し、その有効性と限界を確認した。実験では、学習用データの設定において複数の方式、件名・分類記号ペアの属性群設定において複数の属性集合設定、さらに複数の機械学習法という条件を組み合わせ実験を行った。

(6)国立国会図書館作成の既存書誌・典拠レコードを用いて、著作・表現形データの抽出と統合を試行した。著作データについては研究代表者が以前に試行した実験と同程度に抽出・統合することはできたが、それを超えて網羅的かつ精度高く抽出・統合することはできなかった。また、表現形データの抽出については時間的な制約もあり、報告するに足る成果が得られていない。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 8 件)

谷口祥一, 木村麻衣子. 機械学習によって NDLSH 細目付き件名標目に対する NDC 代表分類記号を同定する試み. 日本図書館情報学会誌, Vol.64, No.2, 2018, 印刷中. 査読有. <https://www.jstage.jst.go.jp/browse/jslis/-char/ja>

Shoichi Taniguchi. Examining BIBFRAME 2.0 from the viewpoint of RDA metadata schema. *Cataloging & Classification Quarterly*, Vol.55, No.6, 2017, p.387-412. 査読有. DOI: <https://doi.org/10.1080/01639374.2017.1322161>

Shoichi Taniguchi. What Does Giving Primacy to a Certain Entity Cause in a Conceptual Model for Cataloging? *Expression-Entity Dominant Model Revisited. Library Resources and Technical Services*, Vol.61, No.4, 2017, p.212-225. 査読有. DOI: <http://dx.doi.org/10.5860/lrts.61n4.212>

谷口祥一, 尾形沙由美. NDLSH における NDC 代表分類記号と書誌レコードの分類記号はどの程度一致するのか. *Library and Information Science*, No.75, 2016, p.37-66. 査読有. <http://lis.msllis.jp/article/LIS075037>

谷口祥一. 図書館目録における「記述」の現状と将来. *アーカイブズ学研究*, No.25, 2016, p.40-55. 査読無.

Shoichi Taniguchi. Modeling resource description tasks in RDA. *Cataloging & Classification Quarterly*, Vol.53, No.1, 2015, p.88-111. 査読有. DOI: <https://doi.org/10.1080/01639374.2014.935544>

谷口祥一. RDA(Resource Description and Access)とその先. TP&D フォーラムシリーズ: 整理技術・情報管理等研究論集, No.24, 2015, p.3-28. 査読無.

谷口祥一. BIBFRAME とその問題点: RDA メタデータの観点から. *情報管理*, Vol.58, No.1, 2015, p.20-27. 査読無. DOI: <http://dx.doi.org/10.1241/johokanri.58.20>

〔学会発表〕(計 9 件)

谷口祥一, 木村麻衣子. 機械学習によって NDLSH 細目付き件名標目に対する NDC 代表分類記号を同定する試み. 2017 年度三田図書館・情報学会研究大会, 2017.

谷口祥一. BIBFRAME 2.0 の意義を問い直す. 2016 年度三田図書館・情報学会研究大会,

2016.

谷口祥一, 木村麻衣子. NDLSH の細目付き件名標目に対して代表分類記号を機械学習によって付与できるか. 第 64 回日本図書館情報学会研究大会, 2016.

谷口祥一. 実体「表現形」優先モデル再論: 特定の実体を優先することの帰結とは何か. 2016 年度日本図書館情報学会春季研究集会, 2016.

谷口祥一. 図書館目録における「記述」の現状と将来. 日本アーカイブズ学会 2016 年度大会, 2016.

谷口祥一, 尾形沙由美. NDLSH における NDC 代表分類記号と書誌レコードの分類記号はどの程度一致するのか. 第 63 回日本図書館情報学会研究大会, 2015.

谷口祥一. RDA モデルの再構築. 第 62 回日本図書館情報学会研究大会, 2014.

谷口祥一. RDA 記述処理タスクのモデリング. 2014 年度三田図書館・情報学会研究大会, 2014.

谷口祥一. RDA(Resource Description and Access)とその先. TP&D フォーラム 2014, 2014.

〔図書〕(計 2 件)

日本図書館情報学会研究委員会編; 宮田洋輔, 谷口祥一, 他 9 名. メタデータとウェブサービス. 勉誠出版, 2016, 207p (p. 23-39)

谷口祥一 緑川信之. 知識資源のメタデータ. 第 2 版. 勁草書房, 2016, 280p.

〔その他〕

ホームページ等

<http://user.keio.ac.jp/~taniguchi/>

6. 研究組織

(1) 研究代表者

谷口 祥一 (TANIGUCHI, Shoichi)

慶應義塾大学・文学部・教授

研究者番号: 50207180

(2) 研究分担者

()

研究者番号:

(3) 連携研究者

()

研究者番号:

(4) 研究協力者

()