

令和 6 年 6 月 6 日現在

機関番号：12102

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K12579

研究課題名（和文）メタデータの利活用性向上を目的としたメタデータプロファイル生成支援と流通基盤構築

研究課題名（英文）Metadata profiles based usability and accessibility enhancements for LOD datasets

研究代表者

永森 光晴（Nagamori, Mitsuharu）

筑波大学・図書館情報メディア系・講師

研究者番号：60272209

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：メタデータ作成者は、その領域の専門家であって、メタデータスキーマ設計に関する経験や知見を持っているとは限らず、メタデータ語彙定義を参照するだけでは、それら語彙の中で定義しているプロパティやクラスといったメタデータタームをLODデータセットの中で適切に組み合わせて構造化することは難しい。そこで本研究では、LODデータセットの利活用を高めるために、適切なメタデータターム選択と利活用性の高いメタデータモデルの設計を支援する必要があると考え、以下の2つの目標を設けて研究を進めた。

- 1) ドメインに応じたメタデータターム選択支援モデルの提案
- 2) FAIR原則を基礎としたメタデータモデル評価支援モデルの提案

研究成果の学術的意義や社会的意義

LODデータセットの作成では、コミュニティの目的に特化した独自のメタデータ語彙に加えて、既存のメタデータ語彙を組み合わせて利用している。しかしながら、LODデータセットのドメインに適切なメタデータ記述項目を表現するための適切なプロパティやクラスといったメタデータタームの選択と、それらメタデータタームを組み合わせてメタデータ記述のための制約を与えた構造の作成は、作成者の知識や経験に依るところが大きい。本研究では、LODデータセットの利活用性向上を目的として、ドメインに適切なメタデータターム選択手法とメタデータモデルの評価手法を明らかにし、そのための支援環境の構築をおこなった。

研究成果の概要（英文）：Metadata creators are experts in their domain and may not have experience or knowledge of metadata schema design, and it is difficult to combine metadata terms such as properties and classes defined in metadata vocabulary definitions appropriately in LOD datasets simply by referring to those vocabulary definitions. It is difficult to adequately combine and structure metadata terms such as properties and classes defined in these vocabularies in an LOD dataset. Therefore, in order to increase the use of LOD datasets, this study considers it necessary to support the selection of appropriate metadata terms and the design of highly usable metadata models, and has set the following two research objectives

- 1) To propose a domain-specific metadata term selection support model.
- 2) To propose a metadata model evaluation support model based on the FAIR principle.

研究分野：情報科学

キーワード：メタデータ メタデータスキーマ セマンティックWeb Linked Open Data

## 1. 研究開始当初の背景

メタデータは「データに関するデータ」と定義され、情報資源の組織化や検索のためだけではなく、保存や流通のための重要な役割を担うようになってきた。現在では、学会、博物館、美術館、政府、地方自治体といった様々なコミュニティが目的に合わせたメタデータを作成し公開している。メタデータの記述方法の定義をメタデータスキーマと呼ぶ。メタデータスキーマは大きく二つに分けられる。ひとつは、Dublin Core や Schema.org に代表されるメタデータの記述に用いる属性語彙（例えばタイトル、作者、更新日など）と属性値語彙（例えば件名標目や分類番号）を定義するメタデータ語彙である。もうひとつは、メタデータ作成の際に、コミュニティが目的に合わせて既存のメタデータ語彙を組み合わせて設計するメタデータの構造や制約（記述項目が必須、省略可能、繰り返しなど）を定義したものである（本研究ではメタデータプロファイルと呼ぶ）。

2001年頃から始まったセマンティック Web の活動によってメタデータの利用はさらに広がっている。これまでのセマンティック Web では、主にメタデータ語彙定義やメタデータの記述に用いる語彙（例えば、RDF Schema, OWL, SKOS など）の標準化がおこなわれてきた。さらに2007年頃からメタデータの利用に焦点を当てた W3C による LOD プロジェクトが開始された。LOD プロジェクトはネットワーク上の RDF 形式で書かれたメタデータを結びつけ、メタデータの相互利用性や流通性を高めることを目的としている。この結びつけられたデータセットを「Linked Data（リンクするデータ）」と呼び、その中でもオープンなライセンスで公開されているものを LOD（Linked Open Data）データセットと呼ぶ。2013年の G8 ロックアーンサミットにおけるオープンデータ憲章合意以降、公開される LOD データセットの数は増加の一途をたどっており、LOD データセットの発見やその利活用性向上が課題となっている。

しかしながら、LOD データセットごとに決まる複数のメタデータ語彙を組み合わせたメタデータの構造（以下、メタデータプロファイルと呼ぶ）の公開は進んでいない。そのため、LOD データセットの発見や理解は手間のかかる作業となっている。そこで本課題では、LOD データセットの利活用性向上を目的として、LOD データセット公開者のメタデータプロファイル生成支援する仕組みと、利用者の LOD データセット利活用支援環境の構築をおこなった。

## 2. 研究の目的

メタデータ語彙の定義は、人が理解するための文書として提供されるだけではなく、RDF Schema・OWL・SKOS（Simple Knowledge Organization System）といった機械が解釈可能な形式で定義される。メタデータ語彙に特化した検索サービスを提供する Linked Open Vocabularies（<https://lov.linkeddata.es/dataset/lov/>）には、2020年10月時点で723件のメタデータ語彙が登録されている。一方、データセットごとに決まるメタデータプロファイルの定義方法についてはいくつかの提案がなされている。Dublin Core Metadata Initiative（DCMI）では、シンガポールフレームワークの中で5つの要素（機能要件、ドメインモデル、記述セットプロファイル、構文ガイドライン、利用ガイドライン）から構成されるメタデータプロファイルを提唱している。また、行政サービスの相互利用性向上を目的とした IMI では DMD（Data Model Description、<https://imi.go.jp/goi/datamodel-about/>）と呼ぶ、データモデルを定義するための仕組みを提案している。しかしながら、データセット公開に合わせてメタデータプロファイルが公開されていることは少ない。公開されていたとしても、人のための文書として公開されているのみで、機械が解釈可能な形式で提供されていない。本研究では、LOD データセットの利活用性を高めるために、メタデータ語彙に加えて、メタデータプロファイルの定義と流通に係る、以下の2つの目標を設けて研究を進めた。

### (1) DCMI シンガポールフレームワークを基礎としたメタデータプロファイル作成支援モデルの

**提案：**メタデータ作成者側にはメタデータの構造定義やメタデータ作成のための人を対象としたガイドラインが作成されているが、機械が解釈可能ではなく、また一般に公開されることは少ない。そこで本研究では、DCMI シンガポールフレームワークを基礎としたメタデータプロファイル作成支援モデルを提案する。本モデルでは、メタデータの構造定義作成に加えて、メタデータを利活用するために必要なデータモデルと利活用のためのガイドラインの作成支援手法の提案をおこなう。

### (2) メタデータプロファイルを基礎とする LOD データセット利活用支援モデルの提案:

LOD データセットの利用者は、目的に合うデータセットを探索し、そしてデータセットの構造を理解して利用する。探索の段階においても、構造理解においても、データセットそのものを参照しなければならず、非常に手間のかかる作業となっている。本研究では、利用者のデータセット利活用性向上を目的として、メタデータプロファイルを利用した LOD データセットの探索支援モデルと、構造理解支援モデルを提案する。

### 3. 研究の方法

まず令和3年度において、メタデータプロファイルを基礎とした LOD データセット利活用支援モデルの検討をおこなった。そして、既存のメタデータ作成プロジェクトを調査しメタデータ提供者側が持つ情報の収集と、データセット利用者のデータセットの探索や比較に関する要求の分析をおこなった。その後、メタデータ提供者のメタデータプロファイルを生成支援機能と、メタデータプロファイルを利用した LOD データセットの構造情報提示機能の2つを持つプロトタイプを作成し、評価をおこなった。

令和4年度は、令和3年度での評価に基づき、メタデータプロファイル作成支援手法の改善を図った。そして、LOD データセットの探索から利用までの過程を支援するためのモデルの提案と、そのためのシステムを構築し、その評価をおこなった。

令和5年度は、前年度までの評価に基づいて、メタデータプロファイル作成支援手法と LOD データセット利活用支援モデルの改善をおこなった。

### 4. 研究成果

#### (1) DCMI シンガポールフレームワークを基礎としたメタデータプロファイル作成支援モデル

2010年代前半からウェブ上でのメタデータ公開が様々なコミュニティによって行われるようになってきた。公開されるメタデータを再利用する、あるいはその構造に従って新規にメタデータを作成するには、メタデータ記述の仕様や文脈に関する情報の整備が不可欠である。再利用性・相互運用性の向上を目的とした、メタデータの仕様書・設計図をメタデータプロファイルと呼ぶ。メタデータの理解に有用なメタデータプロファイルの公開を促進するにはメタデータプロファイル設計・公開のためのガイドラインの共有が求められるが、既存の標準では設計・公開における具体的な要件が示されていない。メタデータ作成者側にはメタデータの構造定義やメタデータ作成のための人を対象としたガイドラインが作成されているが、機械が解釈可能ではなく、また一般に公開されることは少ない。そこで本研究では、メタデータプロファイルの設計・公開を支援するガイドラインを提案することを目的として、過去10年間にウェブ上で公開されたデータセットを対象としたメタデータプロファイルの事例調査を行った。調査内容として、10年前の先行研究の調査との比較検討を行い、メタデータプロファイルを作成する機関や分野の変化や、利用される語彙の傾向の10年前との共通点と相違点を明らかにした。次に、既存のメタデータプロファイル設計の枠組みである DCMI シンガポールフレームワークに基づいた内容調査を行い、共通して書かれる内容や、その記述方式のばらつきを分析した。そして、データ公開のための条項を定める FAIR 原則に基づいたメタデータプロファイルの品質調査を実施し、FAIR 原則を考慮したメタデータプロファイルの設計は十分に行われていないことを示した。さらに、先行研究で行われなかった個々のメタデータチームの利用状況について調査し、参照方法の曖昧さや語彙定義に準拠しない再利用といった課題が明らかにした。

そして、分析結果から得られたメタデータプロファイルの記述内容の傾向と設計・公開における課題に基づいて、メタデータプロファイル構築において記述すべき内容や実施すべき行動を具体化し、メタデータ仕様が明らかになっていることをメタデータ作成者が確認するための指針と作成支援モデルの提案を行った。

#### (2) メタデータプロファイルを基礎とする LOD データセット利活用支援モデル

LOD データセットは、異なる分野のデータセットを組み合わせて利用することが想定されている。しかしながら、複数の LOD データセットから横断的にデータを抽出するための技術である Federated Query を利用した SPARQL 問い合わせでは、しばしば結果が取得できない場合がある。この問題は、問い合わせ先の LOD データセットを公開しているサーバである SPARQL エンドポイントに対して膨大な数の HTTP リクエストが発行されることが原因である。VALUES 句を利用した Federated Query の書き換えによって、発行される HTTP リクエストの数を減らすことが可能だが、書き換え後のクエリは公開された SPARQL エンドポイントごとに設けられた様々な制限により、正常に実行できず結果が取得できない場合がある。

そこで本課題では、LOD データセットから効率的な結果取得を目的として、メタデータプロファイルに基づいて SPARQL エンドポイントの制限を事前に取得し、VALUES 句を利用して Federated Query を書き換えたクエリを事前取得した制限を満たすような粒度に分解して実行する手法を提案した。提案手法の評価のためにいくつかの評価用クエリを用意し、比較手法として Federated Query をそのまま実行する手法と、VALUES 句を利用した書き換え後のクエリをそのまま実行する手法を用意した。結果として、提案手法を利用した評価用クエリの実行は、上述の2つの手法と比較して最速の平均実行時間を記録した。また、提案手法で全ての評価クエリの結果が取得できたわけではないが、最も高い成功率を確認できた。少なからず存在した提案手法でも結果を取得できなかった場合に関して、本研究で定義した公開 SPARQL エンドポイントで設定されている制限以外の種類の制限の存在が直接的な原因であったため、追加の評価用クエリによる実験を行うことで、未定義の制限を新規に発見し、提案手法の改善に繋がると考える。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Thalhath Nishad, Nagamori Mitsuharu, Sakaguchi Tetsuo	4. 巻 1
2. 論文標題 Metadata application profile as a mechanism for semantic interoperability in FAIR and open data publishing	5. 発行年 2024年
3. 雑誌名 Data and Information Management	6. 最初と最後の頁 100068 ~ 100068
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.dim.2024.100068	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 4件）

1. 発表者名 Teruhisa Hanzawa, Mitsuharu Nagamori
2. 発表標題 Ensuring Long-term Access to Metadata Schemas: A Method and Platform for Maintaining Metadata Schemas with a Decentralized Network
3. 学会等名 Proceedings of the 18th International Conference, iConference 2023（国際学会）
4. 発表年 2023年

1. 発表者名 Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi
2. 発表標題 YAMAML: An Application Profile Based Lightweight RDF Mapping Language
3. 学会等名 International Conference on Asian Digital Libraries ICADL 2022（国際学会）
4. 発表年 2022年

1. 発表者名 Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi
2. 発表標題 Application Profile Driven Data Acquisition for Knowledge Graph and Linked Data Generation in Crowdsourced Data Journalism
3. 学会等名 The twentieth International Conference on Dublin Core and Metadata Applications (DCMI 2022)（国際学会）
4. 発表年 2022年

1. 発表者名 岩波春香、永森光晴
2. 発表標題 LODデータセットの構造情報を利用したプロパティ探索支援手法
3. 学会等名 人工知能学会第56回セマンティックウェブとオントロジー研究会
4. 発表年 2022年

1. 発表者名 半澤輝尚、永森光晴
2. 発表標題 メタデータスキーマへの長期的アクセス支援基盤の構築 -分散型ネットワークを用いたスキーマ管理-
3. 学会等名 人工知能学会第56回セマンティックウェブとオントロジー研究会
4. 発表年 2022年

1. 発表者名 Soichiro Watanabe, Mitsuharu Nagamori
2. 発表標題 Federated Query Optimization Method based on SPARQL Endpoint Features for Efficient Retrieval
3. 学会等名 17th International Conference on Metadata and Semantics Research (MTRS2023) (国際学会)
4. 発表年 2023年

1. 発表者名 入江真弘、永森光晴
2. 発表標題 LLMを用いたドメインに基づくメタデータタームと シソーラスのマッピング手法
3. 学会等名 人工知能学会 第62回セマンティックウェブとオントロジー研究会
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------