

氏名	李 春秋
学位の種類	博士（情報学）
学位記番号	博甲第 8744 号
学位授与年月日	平成 30年 3月 23日
学位授与の要件	学位規則第4条第1項該当
審査研究科	図書館情報メディア研究科
学位論文題目	

A Study on Provenance Description of Metadata Schemas for Longevity of Metadata in Networked Information Environment

(ネットワーク情報環境におけるメタデータの長期利用性向上のためのメタデータスキーマの来歴記述に関する研究)

主査	筑波大学	教授	工学博士	杉本 重雄
副査	筑波大学	教授	博士（工学）	森嶋 厚行
副査	筑波大学	准教授	博士（工学）	阪口 哲男
副査	筑波大学	教授	博士（工学）	井上 智雄
副査	京都大学	教授	医学博士	原 正一郎

論文の要旨 (2,000 字程度)

デジタル情報資源（デジタルリソース）の長期保存・長期利用は、ネットワーク情報環境の発展にとって重要であると同時に技術の進歩と広がりゆえに解決が難しい問題であることが広く認められている。特に、World Wide Web (WWW) 環境におけるデジタルリソースの長期利用性を高めることは、WWW が真に我々の社会の情報基盤となるためには必須の課題である。本学位論文はこうした背景の下、デジタルリソースの長期利用性を高める鍵となるメタデータに注目し、メタデータの長期利用性・長期保存性を高める観点から、メタデータの語彙や構造的性質を定義するメタデータスキーマの来歴記述 (provenance description) のためのモデルを提案し、その有効性について論じたものである。以下、本学位論文の構成にしたがって、論文要旨を示す。

第1章から第3章までは、本学位論文の導入と背景説明、並びに意義付けのための関連研究に関する説明がなされている。

第1章は、本論文が述べる研究の背景として、デジタルリソースの長期保存・長期利用に関するこれまでの研究開発プロジェクトとそうしたプロジェクトに関連して開発されてきたデジタルリソースの保存のためのメタデータの標準規格に加え、Dublin Core Application Profile (DCAP) など、本研究の基盤となる WWW 指向のメタデータスキーマの基本モデルを紹介している。

第2章は、DCAP や World Wide Web コンソーシアム (W3C) の PROV 等、本研究の基盤となるモデルと関連する基礎的概念の説明、メタデータの長期利用性に対するメタデータの時間的相互運用性と維持管理の視点からの説明をしたのち、本研究の目的とその意義付けを示している。

第3章は、関連研究のサーベイであり、メタデータの維持管理、メタデータ語彙の維持管理、来歴記述とその標準、図書館等のアーカイブ機関における来歴情報の利用、WWWにおける来歴記述といった領域に分けて論じている。

第4章から第6章は、著者による具体的な研究成果について述べている章である。本研究は、W3Cが進めるLinked Open Data (LOD) の取り組みに適合したメタデータスキーマの来歴記述を行うためのモデルを作ることに加え、DCAP が定義するメタデータスキーマの基本的枠組みに従って、応用ごとに定義されるメタデータの構造的制約とメタデータの記述項目等の語句の意味を定義するメタデータ語彙 (Metadata Vocabulary) を明確に区別して、それぞれの来歴情報を記述することを基本方針としている。そのため、第4章から第6章においては、この基本方針に従って進めた研究内容を示している。

第4章は、本学位論文の核となる研究に先立って行った来歴記述モデルに関する研究について述べている。この研究では、デジタル保存の国際標準メタデータである PREMIS と W3C が開発した来歴記述のモデル PROV で定義された語彙を基礎にして、さらに WWW におけるメタデータの標準モデルである Resource Description Framework (RDF) 上に定義した来歴記述モデルを示している。

第5章は、メタデータスキーマにおける構造制約を決める記述のための来歴記述モデルを定義している。メタデータスキーマにおける構造制約に関し、本研究では DCAP の Singapore Framework が定義する Description Set Profile (DSP) の概念を基盤モデルとして利用し、DSP の来歴記述のためのデータモデルを W3C PROV に基づいて定義している。このモデルは、できるだけシンプルな構造を持つように定義されており、DSP の構成要素である Description Template、Statement Template および Structural Constraints に対して、それぞれ revision、addition、deletion の3種類の activity を定義している。本章では、Digital Public Library of America (DPLA) が定義するメタデータスキーマの3つの連続するバージョンに対して本モデルを適用して RDF によるフォーマルな記述を作成し、それと DPLA のメタデータスキーマ定義の自然言語 (英語) 記述による変更記録とを比較評価することで、本研究による提案モデルの優位性を示している。

第6章は、メタデータ語彙に関する来歴記述モデルを示している。第5章で示したモデルと同様に、W3C の PROV を基礎として、メタデータ語彙に含まれるメタデータタームの変更を記録するためのシンプルなデータモデルと、そのモデルのための RDF によるフォーマルな記述方式を定義している。本章では、ここで定義したモデルを LCSH のターム等に適用した例を示している。

第7章及び第8章は本学位論文の結びにあたる部分である。デジタルリソースの長期利用・長期保存に関する研究は1990年代から進められてきているが、メタデータとそのスキーマをデジタルリソースとしてとらえ、かつそれらを WWW 上で蓄積共有されるリソースとしてとらえて来歴記述を行うことを目指した研究は多くない。結びの部分では、こうした視点も含めた考察が示されている。具体的には、第7章では、最初にメタデータの長期保存と一般のデジタルリソースの長期保存の違い、メタデータの長期保存のための重要な視点、そして WWW 環境における来歴記述の要件といった視点からの考察を示し、その後で、研究成果の範囲に対する検討、将来に向けた課題等について述べている。第8章は、結びの章として、研究活動全体を通じて著者が得た知見や将来に向けた検討課題等について述べている。

審査の要旨 (2,000 字以上)

【批評】

論文要旨に示したように、本論文は、メタデータの長期利用性を高める観点から、メタデータの語彙や構造的性質を定義するメタデータスキーマの来歴記述に注目し、Linked Open Data 環境におけるメタデータ交換のための標準である Resource Description Framework (RDF)、World Wide Web コンソーシアム (W3C) によって開発された来歴記述標準である PROV、さらにインターネット上のメタデータ相互運用性のための基盤的モデルである Dublin Core Application Profile 等、広く認められた基盤的モデルの上に、できるだけシンプルなメタデータスキーマの来歴記述モデルを提案し、その有用性を論じたものである。インターネットの爆発的な広がりやデジタルアーカイブ開発の進展の中で、デジタルリソースの長期保存や長期利用に関する研究開発や標準規格の開発は進んできた。そうした研究開発の中で、メタデータがデジタルリソースの長期保存・長期利用に重要な役割を果たすことは認められてきたが、その一方、メタデータそのものの長期利用性を支えるメタデータスキーマの長期利用・長期保存に関する議論はあまりなされてこなかった。これは、メタデータスキーマが文書として作られ、スキーマの保存が文書の保存に置き換えられていたためであると考えられる。しかしながら、WWW の発展によりメタデータスキーマ自体も、ある種のメタデータとして WWW 上で送受信、共有されるデジタルリソースになってきているため、従来の文書としてのスキーマ保存のモデルだけでは対処できなくなっている。本論文が述べる研究は、こうした現状と将来へのニーズに関する理解の下に進められたものである。以下、本論文の各章に関する批評を述べ、その後にもとめを示す。

第 1 章、第 2 章は導入部として、研究背景から始め、研究の位置づけ、基本概念、ならびに研究目的等について述べている。第 1 章では、デジタルリソースとメタデータの長期保存・長期利用の間の関連性とその違いなど、本論文が述べる研究の位置づけを理解するための背景について述べている。それに続き、第 2 章は、はじめに、メタデータの長期利用性、来歴記述やメタデータの相互運用性のためのモデルに関して述べ、時間的な相互運用性の視点からメタデータの長期利用性について示している。その後、本研究の目的と新奇性に関して論じている。第 1 章、第 2 章は本学位論文の導入部分として十分な内容をカバーしていると思われる。第 2 章と第 3 章との重なりは少しあるが、本論文としての価値を損なうものではない。

第 3 章は、関連研究に関するレビューを示している。この章では、メタデータのキュレーションと管理、メタデータ語彙のメンテナンス、メタデータレジストリ、来歴情報の記述、関連する標準規格、メタデータの変更履歴追跡、アーカイブ機関における来歴記述の利用、WWW における来歴記述といった視点から関連する研究や事例について述べている。WWW 環境向けのメタデータスキーマの長期利用・長期保存が新しい研究領域であるため、直接的に本研究の比較の対象となるような研究は関連研究として含まれていないが、本研究で開発したモデルの新奇性を理解する上では十分な範囲が含まれていると思われる。

第 4 章から第 6 章までは本論文が提案するモデルについて述べている。第 4 章は、W3C によって開発された来歴記述モデル PROV とデジタルリソースの長期保存のためのメタデータの国際標準 PREMIS で定義された概念の構造を組み合わせ、RDF に基づく WWW 指向のメタデータスキーマ

マの来歴記述モデルの基盤を与えている。この章自体は、モデルの提案と記述方法の定義にとどまっており、提案モデルの記述能力の評価等には踏み込んでいない。しかしながら、第5章と6章に示されるメタデータスキーマの構造制約とメタデータ語彙の来歴記述のモデルの基盤としては十分な内容を含んでいると理解できる。

第5章は、Dublin Core Metadata Initiative が提案した Singapore Framework の Description Set Profile の概念に基づいてメタデータスキーマの構造記述をとらえ、その来歴記述モデルを提案し、Digital Public Library of America の Description Set Profile 定義の連続するバージョン間での変更履歴記述に対して提案モデルを適用することで、提案モデルの持つ有用性を示している。提案モデルはシンプルであるため、その記述性や適用範囲の検証が課題である。この課題に対して、本研究は、メタデータスキーマを RDF 記述し、実際に SPARQL エンドポイントとして利用できる形にしたうえで、評価に利用することで実用性を示している。WWW 上で公開されているメタデータスキーマの変更履歴が限られることもあり、評価例は1つしかないが、評価例の選択基準や SPARQL を利用した評価が十分に述べられており、研究としての信頼性は損なわれていないと考えられる。

第6章が示すメタデータ語彙の来歴記述モデルは、第5章で示したモデルと同様にできるだけシンプルな構造を持つように定義されている。ここでは、メタデータ語彙とその要素であるタームの変更パターンをいくつかに分けて定義し、例を用いてその具体的な記述を示している。モデルの評価に関する記述が少ない点がやや気にはなるが、モデル記述そのものはフォーマルになされており、研究としての評価は十分に可能である。

第7章は、本研究の寄与する範囲や限界等に関して全体的な考察を述べており、第8章は本論文の結びを述べている。これらの章は、本研究の遂行によって得られた知見と経験を適切にまとめている。

本論文全体として、ネットワーク情報化時代における重要な課題として広く知られているデジタルリソースの長期利用の問題に対して、従来はあまり広く認知されていなかったメタデータスキーマの長期利用・長期保存の問題に注目し、来歴記述という視点と WWW 上でのメタデータの相互運用性の視点からアプローチして得た価値ある研究成果を適切にまとめた学位論文であると考えられる。

【最終試験結果】

平成30年1月31日、図書館情報メディア研究科学学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程（課程博士）の学位論文審査に関する内規」第23項第3号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

【結論】

よって、著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。