

## 2.5 学位論文のメタデータ

### 2.5.1 はじめに

本学をはじめとした科学技術系の大学院においては修士および博士の学位を取得するためには、学位論文を作成し、審査を受け、一定の水準に達していると判断される必要がある。本学は3つの研究科から構成され、年間およそ350名程度が修士の学位を取得するため論文を作成し、修士号が授与されている。博士の学位については、およそ60名程度が論文を作成し、博士号が授与されている。学生が学位取得のために作成した論文は、修士号の場合は修士論文、博士号の場合は博士論文と呼ばれている。また本学には学部がないが、学部を卒業する際には論文を作成し、審査を受け、一定の水準に到達していると判断された場合は学士号が授与される。この場合は通常卒業論文と呼ばれている。以降、卒業論文、修士論文、博士論文を総称して学位論文と呼ぶ。

博士論文の場合は、提出された論文は、大学附属図書館、研究科および国立国会図書館において保管され、公開されている。国立国会図書館において保管、公開されることにより、その内容が一般に公開されることが保証されている。修士論文、卒業論文の場合には、その保管方法は大学にゆだねられており、多くの場合は積極的な公開がなされていない。

学位論文には、学生が在学中に行った研究内容が詳細に述べられている。学生が在学中に行った研究を報告する場としては、学術論文誌への論文投稿、国際会議などの会議などがある。学術論文誌に掲載された場合はなんらかの形でその存在が知られ、研究成果が流通する。しかし学位論文の場合は関係者以外に知られることは少ない。このような状況にある学位論文の流通させるためには、その内容を含めた存在情報を発信する必要がある。つまり、学位論文の情報 - メタデータ - を整理し、ネットワークに発信してゆくことが重要になる。

メタデータについては国際的な標準として Dublin Core が提案されている。現在、この Dublin Core に準拠して、様々な情報についてのメタデータの詳細が議論されている。そこで、この Dublin Core に準拠した形で学位論文のメタデータの構造を定義し、付与することを考える。メタデータの構造を考える際には、国際的な情報流通を可能にするために、海外で行われている同種のプロジェクトを参考とし、将来的には参加することが可能なようにすることが望ましい。そこで、次節では米国を中心として行われている同種のプロジェクトについてのべる。

## 2.5.2 米国での学位論文データベースプロジェクト

米国での学位論文を集積しようというプロジェクトが行われている。そのプロジェクトについて述べる前に、米国での学位論文の取扱いの現状について概観する。

米国においても、我が国と同様に修士、博士の学位を取得するには論文を提出し、審査を受ける必要がある。提出された論文が要件をみたしていると判断されれば、修士もしくは博士の学位が授与される。ただ、提出された学位論文の取扱いには特にきめられたものがあるわけではない。学位申請をした学生が、自らの判断で出版社（UMI社であると聞く）に論文を送付する。送付を受けた出版社は、学位論文をマイクロフィルム化し、出版する。ここで注意しなければならないのは、学位論文を出版社に送付することは学生の義務ではなく、自由意志によるものであるという点である。あるいは、学位論文を学術図書として出版する。これは修士論文、博士論文ともに同じような状況であり、博士論文を国立国会図書館に送本するシステムが確立された我が国とは大きく異なる。このため、米国では我が国よりも学位論文の流通がシステム化されていない一面がある。

近年のインターネットに代表される情報ネットワークの発達により、学位論文を収集し、それをデータベース化することが比較的容易に実現できる可能性が出てきた。これを実現しようとして、NDLTDとよばれるプロジェクトが始められた。NDLTDとはNetworked Digital Library of Thesis and Dissertationsの略であり、そのホームページは<http://www.ndltd.org/>である。バージニア工科大学のEdward Fox教授が中心となって、米国のみならず世界の大学に呼び掛け、学位論文を電子化し、一種の電子図書館として実現しようとするものである。ネットワーク上に構築されることから、国境を越えた学位論文の集積を行うことも目的の一つになっており、メキシコ、シンガポール、韓国などの大学が参加している。ただし、日本から参加している大学はまだない。

このNDLTDでは、学位論文を蓄積する際に学位論文のメタデータを定義している。Dublin Coreの基本的な考え方は、情報（この場合は学位論文）の作成者が同時にメタデータも作成する。そうすることによって、メタデータを付与する専門家が必要になることを防ぎ、情報の公開と同時にメタデータが利用できるようにしようというものである。この考えにそって、NDLTDでは論文作成を行う学生に対して、メタデータを入力するようなガイドラインを設けている。NDLTDで規定されているメタデータを次に示す。

DC.Title	必須：学位論文のタイトル
DC.Title.X-Notes	任意：タイトルに関連するメモ
DC.Creator.PersonalName	必須：著者の名前(姓, 名の順 . BibTex に準拠)
DC.Creator.Address	任意：登録時の著者の住所
DC.Creator.X-Institution	必須：登録時の著者の所属
DC.Creator.X-Major	推奨：著者が所属した専攻
DC.Creator.X-College	任意：著者が所属した学部
DC.Creator.X-Dept	推奨：著者が所属した学科
DC.Subject	任意:(いわゆる)論文のキーワード
DC.Description	対象とするデータが画像や図表などの場合のキャプション
DC.Description.Abstract	推奨：論文の梗概
DC.Publisher.CorporateName	推奨：学位論文を電子的に提供する際の責任組織の名前
DC.Publisher.CorporateName.Address	任意：学位論文を電子的に提供する際の責任組織の住所 .E-mail アドレス .

DC.Contributor.X-Chair.PersonalName	推奨：主査の名前
DC.Contributor.X-Chair.PersonalName.Address	任意：主査のアドレス
DC.Contributor.X-Committee.PersonalName	推奨：副査の名前
DC.Contributor.X-Committee.PersonalName.Address	任意：副査のアドレス
DC.Date.Valid	必須：卒業年次．4桁 の西暦で記述
DC.Date.X-Defence	任意：審査が行われた 年
DC.Type	必須：Dublin Core で 定義されるタイ プ．博士論文は Text.Thesis.Do ctoral , 修士論 文は Text.Thesis.Ma sters , 学士論文 は Text.Thesis.Ba chelors .
DC.Format	必須：IMT (Internet Media Type, MINE)
DC.Format.X-ContentLength	任意：論文データの長 さ(バイト単位)
DC.Format.X-Checksum	任意：論文データの MD5 によるチェ ックサム
DC.Identifier	必須：このメタデー タが示す論文本体 の URN , URL
DC.Identifier.X-CallNumber.LC	任意：LC の登録番号
DC.Identifier.X-CallNumber.Local	任意：Local に与えら れた登録番号

DC.Source	任意:この論文の出所 . スキャンされた 画像の場合は出 典のこと
DC.Language	任意:論文の記述言語
DC.Relation	必須:論文に関連する URL , URN
DC.Coverage	任意:論文が記述する 内容の物理的な もの , あるいは 時間的なもの . DC-wp-3
DC.Rights	任意:論文の権利関係 の記述 (ポイン タもあり)
DC.Rights.X-Availability	必須:論文の閲覧範囲 . 次のものから選 択する . worldwide, university, campus, college, secure
DC.Rights.X-Availability.Notify	任意:閲覧範囲が有効 となる日付け
DC.Rights.X-Proxy.PersonalName	任意:著者に連絡がと れない時に著者 の代理人となる べき人の名前
DC.Rights.X-Proxy.Address	任意:上記代理人のア ドレス
DC.Rights.X-Signature	任意:電子署名 (PGP)

### 2.5.3 奈良先端科学技術大学院大学での学位論文メタデータ付与の検討

前節で述べたように，NDLTD では学位論文に対してのメタデータを定義している．ここでは，これが我が国の学位論文に適用できるかを検討してみる．

我が国で作成される学位論文の多くは日本語で記述されている．NDLTD におけるメタデータは米国（英語圏）の研究者が考案しているため，米国の事情が反映されているが他の国の事情が反映しているとは言えない．具体的な例をあげると，DC.Identifier.X-CallNumber.LC という項目があり，そこには LC(Library of Congress，米国議会図書館)への登録番号を記述するようになっている．米国において出版されるほぼすべての書籍は議会図書館に登録申請が行われ，それに基づきカタログ番号が議会図書館から付与される．様々なカタログ（文献の二次情報のデータベース）や OPAC ではこの議会図書館から付与された登録番号を用いて処理が行われており，米国では学術情報の流通システムにおいて必要不可欠なものとなっている．しかし，米国以外の国において出版される書籍について、米国議会図書館にカタログ番号の登録が請求されることは少なく，特に我が国における出版物についてはほとんど行われていない．学位論文についてはほとんどこのようなことが行われていないものと考えられる．このため，このようなメタデータ項目は我が国では意味を持たない．

また，我が国のように英語以外の言語で学位論文を作成することが多い場合にはメタデータも学位論文が記述されている言語で記述されることが多くなる．例えば，論文の概要（これは大学や学部，研究室などによって梗概，あらましなどとも呼ばれているようである）は，通常論文が記述されている言語が用いられる．我が国の学位論文についてみると，この点は，日本語で記述された学位論文でも最近では英語による概要も同時に作成するよう義務付けている場合が多いが，すべての学位論文で行われているとは限らない．概要だけでなく，論文題目についても同じようなことがいえる．しかし，国際的な流通を考えた場合には英語による記述が欠かせない．この問題に対しては，これから作成される学位論文については英語によるメタデータの記述を推奨することによって対処できるが，既存のものについては別途作成しなければならない．この場合，学位論文作成者が自ら作成するものでないため正確性の問題が残る．学位論文の分野に精通した研究者が作成することで解決するしかない．

先にも述べたように、NDLTD のメタデータは英語を母国語とする研究者が考案しているため、英語以外の多数の言語にそのまま対応できるかという問題点がある。米国をはじめとした研究者の多くの母国語は英語もしくはいわゆる印欧語族の言語である。印欧語族の言語を記述する際には、アルファベットを起源とした文字を用いる。これに対して、日本語、中国語をはじめとするアジア言語では表意文字である漢字を使用したり、それ以外の起源を持つ文字を用いて記述される。この印欧語族の言語とアジアの言語との違いは大きく、例えば文字を計算機内部で表現する際の文字コードにもその問題点が見られる。一般に印欧語族で用いられるアルファベットは1バイトのコードが割り当てられていることが多く、またそれらの文字が混在することに大きな問題は生じない。これに対して、アジアの言語は計算機内部での表現コードが2バイトから3バイトで表現されることが多い。日本語に限ってみても、現在日本で使用されている漢字すべてが文字コードに表現されているわけではなく、文字コード表に存在しない漢字も数多い。特に、人名で用いられる漢字が文字コード表にないことがよく見られる。これらに対応するために、外字というものが定義されているが、利用者が統一性なく文字を割り当てているため、システムごとの互換性がない。地方自治体の戸籍システムではこれが顕著で、もはや全国共通のシステムとすることが不可能であることは有名である。中国語にいたっては、漢字が2種類存在し、それぞれ独自の文字コードを持つ。しかも問題となるのは、アジア系言語の文字を混在させることが難しいことである。どれか一つのアジア言語と英語の同時表示は問題なく行える（標準化されている）が、日本語と中国語を一つの文書内で表現する方法は標準化されていない。これは文字コードの問題ではあるが、現実にはメタデータを国際的に流通させようとした場合には大きな障害となる。現実的にはメタデータを記述する際に、使用する文字コードセットを指定する。そして、メタデータを受け入れるシステムが対応できる場合には使用するが、そうでない場合は棄却するようにすればよい。文字コードの指定方法としてはメタデータの記述に XML などを利用して、メタデータが記述されている言語（文字コード）を示すことが考えられる。

日本語特有の問題としてはふりがなあるいはルビがあげられる。日本語では、氏名や読みの難しい漢字にどう読むのかを示す仮名をつけて、正しい発音を示す。特に氏名の場合は、ふりがながないと読めない場合がままある。学位論文のメタデータには著者である学生の名前を含むため、この問題は避けて通れな

い。この対処法としてはふりがなのためのメタデータ項目を作成することが一つあげられる。この方法では、ふりがなのために別途メタデータに項目を追加することになる。Dublin Core の考え方では、日本語のためにメタデータの項目を定義してもよいという考え方である。日本語が処理できるシステムではそのメタデータを使用すればいいが、日本語を扱えないシステムではこの項目を無視すればよい。ただ、日本語が扱えるのに使用しないシステムが発生する可能性がある。それに日本語のためだけにメタデータの項目を追加するという問題点もある。

もう一つの解決法として、氏名の英文表記からふりがなを生成することが考えられる。通常、氏名の英語名には日本語の発音をローマ字で表記したものをしていることが理由である。これは著者である学生が日本人の場合にはうまく機能すると期待できる。しかし現実には留学生が作成する学位論文も数多くあり、この場合にはうまくいかない。特に漢字文化圏の出身学生の場合に問題となる。非漢字文化圏の学生の場合は、もともと日本語で表記された氏名がふりがなとなっているため問題にならない。英語表記をローマ字とみなしてかなに変換し、問題がなければふりがなとして採用すれば大きな間違いにはならないと予想される。しかし、間違いなく変換できるという保証はない。

このように、NDLTD で定義されている学位論文のメタデータについては、国際的なメタデータの流通を前提とするならば積極的に使用すべきではあるが、細かい点で検討と解決策を考案しなければならない。

#### 2.5.4 学位論文のメタデータ付与への取り組み

前節で示したような問題点がある。本来はこれをメタデータレベルで解決すべきだが、まず学位論文のデータベースを作成するために必要となるデータベースのスキーマの定義について述べる。メタデータとしての項目はそのままデータベースのスキーマとして利用できるようになっているが、今回はスキーマとして設計する際に、日本語と英語のそれぞれについてメタデータ項目をスキーマとして定義した。さらに前節で明らかにしたふりがなの問題にも対応するために、暫定的にふりがなを入力すべきレコードを作成した。定義したスキーマを以下に示す。



DC.Title	必須：学位論文のタイトル（英文）
DC.Title.Y-Native	必須：学位論文のタイトル（和文）
DC.Title.X-Notes	任意：タイトルに関連するメモ（英文）
DC.Title.X-Notes.Y-Native	任意：タイトルに関連するメモ（和文）
DC.Creator.PersonalName	必須：著者の名前（姓，名の順 .BibTex に準拠，英文）
DC.Creator.PersonalName.Y-Native	必須：著者の名前（姓，名の順 .和文）
DC.Creator.Address	任意：登録時の著者の住所（E-mail）
DC.Creator.X-Institution	必須：登録時の著者の所属大学（英文）
DC.Creator.X-Institution.Y-Native	必須：登録時の著者の所属大学（和文）
DC.Creator.X-Major	必須：著者が所属した専攻（英文）
DC.Creator.X-Major.Y-Native	必須：著者が所属した専攻（和文）
DC.Creator.X-College	必須：著者が所属した学部（英文）
DC.Creator.X-College.Y-Native	必須：著者が所属した学部（和文）
DC.Creator.X-Dept	必須：著者が所属した学科（英文）
DC.Creator.X-Dept.Y-Native	必須：著者が所属した学科（和文）
DC.Subject	必須：（いわゆる）論文のキーワード（英

DC.Subject.Y-Native	文) 必須:(いわゆる)論文のキーワード(和文)
DC.Description	対象とするデータが画像や図表などの場合のキャプション(英文)
DC.Description.Y-Native	対象とするデータが画像や図表などの場合のキャプション(和文)
DC.Description.Abstract	推奨:論文の梗概(英文)
DC.Description.Abstract.Y-Native	推奨:論文の梗概(和文)
DC.Publisher.CorporaterName	必須:学位論文を電子的に提供する際の責任組織の名前(英文)
DC.Publisher.CorporaterName.Y-Native	必須:学位論文を電子的に提供する際の責任組織の名前(英文)
DC.Publisher.CorporateName.Address	任意:学位論文を電子的に提供する際の責任組織の住所.E-mailアドレス.
DC.Contributor.X-Chair.PersonalName	推奨:主査の名前(英文)
DC.Contributor.X-Chair.PersonalName.Y-Native	推奨:主査の名前(和文)
DC.Contributor.X-Chair.PersonalName.Address	任意:主査のアドレス(E-mail)
DC.Contributor.X-Committee.PersonalName	推奨:副査の名前(英

DC.Contributor.X-Committee.PersonalName.Y-Native	文) 推奨：副査の名前（和文）
DC.Contributor.X-Committee.PersonalName.Address	任意：副査のアドレス（E-mail）
DC.Date.Valid	必須：卒業年次．4桁の西暦で記述
DC.Date.X-Defence	任意：審査が行われた年．4桁の西暦で記述
DC.Type	必須：Dublin Coreで定義されるタイプ．博士論文は Text.Thesis.Doctoral, 修士論文は Text.Thesis.Masters, 学士論文は Text.Thesis.Bachelors.
DC.Format	必須：IMT (Internet Media Type, MIME)
DC.Format.X-ContentLength	任意：論文データの長さ（バイト単位）
DC.Format.X-Checksum	任意：論文データのMD5によるチェックサム
DC.Identifier	必須：このメタデータが示す論文本体のURN, URL
DC.Identifier.X-CallNumber.LC	任意：LCの登録番号
DC.Identifier.X-CallNumber.Local	任意：Localに与えられた登録番号
DC.Source	任意：この論文の出所．スキャンされた画像

DC.Source.Y-Native	<p>の場合は出典のこと (英文)</p> <p>任意：この論文の出所 . スキャンされた画像 の場合は出典のこと (和文)</p>
DC.Language	必須：論文の記述言語
DC.Language.Y-Native	<p>必須：このメタデータ の Native での記述言 語</p>
DC.Relation	<p>必須：論文に関連する URL , URN</p>
DC.Coverage	<p>任意：論文が記述する 内容の物理的なもの , あるいは時間的なも の . DC-wp-3</p>
DC.Rights	<p>任意：論文の権利関係 の記述 ( ポインタもあ り , 英文 )</p>
DC.Rights.Y-Native	<p>任意：論文の権利関係 の記述 ( ポインタもあ り , 和文 )</p>
DC.Rights.X-Availability	<p>必須：論文の閲覧範囲 . worldwide, university, campus, college, secure があ る</p>
DC.Rights.X-Availability.Notify	<p>任意：閲覧範囲が有効 となる日付け</p>
DC.Rights.X-Proxy.PersonalName	<p>任意：著者に連絡がと れない時に著者の代 理人となるべき人の 名前 ( 英文 )</p>

DC.Rights.X-Proxy.PersonalName.Y-Native	任意：著者に連絡がとれない時に著者の代理人となるべき人の名前（和文）
DC.Rights.X-Proxy.Address	任意：上記代理人のアドレス（E-mail）
DC.Rights.X-Signature	任意：電子署名（PGP）

以上のようなスキーマを定義した .このスキーマを用いてデータベースを構築しようとしている .現在は ,入力用のインタフェースを作成したところである .この入力用インタフェースは WWW 経由で入力作業ができるように設計している .

#### 2.5.5 さいごに

以上 ,学位論文をネットワーク上で公開する場合に必要となってくるメタデータについての検討を行っている現状を報告した .今後実際にメタデータを入力し ,本学で電子化されて公開されている学位論文と関連付けてその有効性の検証を行う必要がある .さらに国際的な学位論文の情報交換が行えるかを ,たとえば NDLTD のようなプロジェクトと共同して検証しなければならない .

日本から欧米諸国に対する学術情報の発信が求められているが ,そのような要求にこたえるためにも ,学位論文のメタデータの整備を進めていく必要があると考えている .