

CA1900

国立国会図書館による 博士論文収集の現況と課題

わたなべ しゅん*
渡部 淳*

1. はじめに

2013年4月1日に、学位規則の一部を改正する省令(平成25年文部科学省令第5号)が施行され、博士の学位の授与に係る論文(以下「博士論文」という)は、インターネットの利用により公表することとされた(E1418参照)。

国立国会図書館(NDL)は、電子形態で公表される博士論文について、学位授与機関からの送信による収集を2014年2月4日に、機関リポジトリからの自動収集を2015年2月1日に開始した。

本稿は、収集開始から3年が経過したのを機に、NDLにおける学位規則改正後の博士論文の収集状況を紹介するとともに、学位授与機関における博士論文のインターネットによる公表状況についての調査結果を報告する。

2. NDLにおける博士論文の収集

2.1. 収集方法

NDLは、学位規則改正後の博士論文について、公表形態に応じて、以下のように収集している。

第一に、所定の条件⁽¹⁾を満たした機関リポジトリで公表される博士論文は、国立情報学研究所(NII)の学術機関リポジトリデータベース(IRDB)と連携し、システムにより自動収集する⁽²⁾。

第二に、学位授与機関のウェブサイト等で公表される博士論文は、当該学位授与機関からの送信により収集する。

第三に、学位規則で定める「やむを得ない事由」があり、全文がインターネット公表されない博士論文については、電子形態であれば当該学位授与機関からの送信により収集し、冊子形態であれば学位規則改正前と同様、冊子形態のものを学位授与機関からの送付により収集する。

なお、文部科学省では、教育研究成果のオープンアクセス(OA)化を含め知的情報の蓄積・発信のための重要な手段として機関リポジトリを位置付けており、機関リポジトリでの博士論文公表を原則としている⁽³⁾。NDLにおいても、学位授与機関からの送信・送付作業が不要となるため、可能な限り機関リポジトリで博士論文が公表されることを期待している。

2.2. 収集状況

NDLにおける学位規則改正後の博士論文の収集実績(2017年3月末現在)は表1の通りである。

表1 博士論文の収集実績(2017年3月末現在)

	収集件数
自動収集	22,229
電子形態の送信	9,042
冊子形態の送付	4,746
合計	36,017

(注) 2013年度以降学位授与分

文部科学省によれば、2013年度の学位授与件数は1万5,427件であるが⁽⁴⁾、2014年度以降の学位授与件数については、まだ公表されていない。そこで、NDLが受領した学位授与報告書の写し⁽⁵⁾を利用して、収集率を試算した。

NDLが2017年3月末までに受領した学位授与報告書の写しの件数は4万4,068件である。よって、収集率は81.7%、未収集件数は8,051件であり、18.3%が収集できていないと推測される。

2.3. 未収集の原因

2016年4月に、収集率が20%未満かつ未収集の論文が20件以上ある34大学の2,950件の博士論文について、未収集の原因を調査した。結果は表2の通りである。

表2 博士論文の未収集の原因

原因	大学数	論文数
(1) 機関リポジトリで博士論文全文を公表しているが、「著者版フラグ」に「ETD」と入力されていない。	7	602
(2) 機関リポジトリで博士論文全文を公表しているが、junii2ガイドラインで定められた書式と異なった記述をしているため、IRDBへのハーベストがエラーになっている。	6	445
(3) 上記(1)(2)の原因が重複している。	11	1,189
(4) 機関リポジトリで博士論文全文を公表しているが、機関リポジトリがIRDBへのハーベスト対象外である。	2	55
(5) 機関リポジトリで博士論文全文を公表しているが、機関リポジトリの仕様により、PDFファイルではなくHTMLファイルが収集されてしまう。	1	159
(6) 機関リポジトリでも大学のウェブサイトでも博士論文全文を公表していない。	7	500
合計	34	2,950

まず、IRDBとの連携により自動収集の対象となるのは、NIIが策定したメタデータ・フォーマット「junii2」⁽⁶⁾に対応している機関リポジトリに登録されている博士論文のうち、以下の2つの条件を満たすものである。

- ・「著者版フラグ(textversion)」⁽⁷⁾の値が「ETD」
- ・「学位授与年月日」が「2013-04-01」以降の日付

*関西館電子図書館課

すなわち、機関リポジトリに博士論文全文を登録していても、「著者版フラグ」の値が「ETD」以外だと、NDLに自動収集されない。

また、junii2ガイドラインで定められた書式と異なった記述をしているため、機関リポジトリのIRDBへのハーベストがエラーになることがある。例えば、junii2バージョン3.1では、学位授与番号は「10100甲第123456号」のように、「科研費機関番号(5桁)+[甲|乙|*]+第*+報告番号+号」(*は0字以上の任意の文字列)の形式で記述するルールとなっている⁽⁸⁾。科研費機関番号を入力せず、「甲第123456号」とのみ記述すると、エラーが発生し、IRDBにハーベストされず、NDLが博士論文を自動収集できない。

NDLでは、未収集の原因を特定後、表2の(1)から(3)の原因を有する大学に対しては、正確なメタデータ処理をするよう依頼している。表2の(4)から(6)の原因を有する大学に対しては、電子形態の博士論文を送信するか、又は郵送により冊子形態の博士論文を送付するよう依頼している。

2.4. 収集すべきでないデータの誤収集

NDLでは、機関リポジトリから自動収集した博士論文について、メタデータ・全文ファイル等を確認した上で、国立国会図書館デジタルコレクション⁽⁹⁾で公開している。しかし、本来収集対象ではないファイルが誤って収集されているケースがある。2017年3月までに誤って収集されたデータは2,108件に達し、自動収集による収集件数全体の8.7%に相当する。

誤収集の主なケースは以下の通りである。

- ・要約又は要旨のみの公表であるが、「著者版フラグ」に「ETD」と入力されている。
- ・別の著者の博士論文が誤って登録されている。
- ・博士論文本文へのリンクURLが誤っているため、PDFファイルではなく、HTMLファイルが収集される。

これら収集対象外のデータについては、国立国会図書館デジタルコレクションでは公開していない。

その他にも、著者名やタイトルに誤字脱字やスペルミスがある、学位授与番号や学位授与年月日が学位授与報告書の記載と一致していない等のメタデータの誤りも散見される。

NDLでは、以上のような状況が判明した場合、学位授与機関に対してメタデータの修正を依頼している。

3. 学位授与機関におけるインターネット公表状況

学位規則の改正によって、博士論文は原則として、

その全文がインターネット公表されることになった。しかし、博士論文が立体形状による表現を含む場合や、全文の公表により学位授与者に明らかな不利益が生じる場合など、「やむを得ない事由」があると学位授与機関が承認した場合には、博士論文の全文に代えて、その要約を公表することができると定められている⁽¹⁰⁾。

機関リポジトリ推進委員会の調査によれば、2013年度学位授与の博士論文について、2014年11月3日時点での全文公表率は約28%であった(E1707参照)。

そこでNDLでは、学位規則改正後の博士論文について、2017年3月末時点での学位授与機関におけるインターネット公表の状況を調査した。

博士論文全件の公表状況を調査するのは困難なため、まず以下のように類型化し、「公表」又は「未公表」と判定した。

- (1) NDLが機関リポジトリから自動収集した博士論文→「公表」と判定。
- (2) NDLが学位授与機関から電子形態の送信により収集した博士論文
 - ・「掲載URL」入力あり→サンプル調査を実施。
 - ・「掲載URL」入力なし→「未公表」と判定。
- (3) NDLが学位授与機関から冊子形態の送付により収集した博士論文→「未公表」と判定。
- (4) NDLが未収集の博士論文→サンプル調査を実施。

(2)について、NDLでは学位授与機関に対して、インターネット公表している博士論文を送信する場合、メタデータ項目の「掲載URL」を入力するよう、依頼している。

ただ、「掲載URL」が入力されていても、博士論文全文ではなく、要約のみ公表の場合があるので、「掲載URL」入力ありの博士論文については、サンプル調査を実施し、全文を公表しているかどうかを調べた。一方、「掲載URL」入力なしの博士論文については、「未公表」と判定した。

電子形態の送信により収集した博士論文は9,042件であり、このうち「掲載URL」の入力があるものは327件であった。この327件のうち無作為に抽出した30件を調査したところ、博士論文全文のインターネット公表が確認できたのは、22件(73.3%)であった。よって、 $327 \times 73.3\% = 240$ 件を「公表」と推計した。

(4)について、NDLが未収集の博士論文でも、NDLが収集できていないだけで、インターネット公表している場合がありうる。そこで、「2.3 未収集の原因」の表2で挙げた34大学の博士論文を用いて、サンプル調査を実施した。34大学2,950件のインターネット公表状況は、表3の通りである。

表3 未収集博士論文のインターネット公表状況

インターネット公表状況	大学数	論文数
機関リポジトリで博士論文全文を公表している。	27	2,450
大学のウェブサイトで博士論文全文を公表している。	0	0
機関リポジトリでも大学のウェブサイトでも博士論文全文を公表していない。	7	500
合計	34	2,950

インターネット公表が確認できたのは、2,450件(83.1%)であった。「2.2 収集状況」で算出した通り、未収集の博士論文は8,051件である。よって、8,051件×83.1%=6,690件を「公表」と推計した。

以上をまとめると、以下の表4の通りである。全体のインターネット公表率は66.2%程度と推測される。

表4 博士論文全文のインターネット公表率

	件数	公表件数	公表率
自動収集	22,229	22,229	100.0%
電子形態の送信	9,042	240	2.7%
冊子形態の送付	4,746	0	0.0%
未収集	8,051	6,690	83.1%
合計	44,068	29,159	66.2%

(注)「電子形態の送信」「未収集」の公表件数については推計値

おわりに

学位規則の改正によって、原則として、すべての博士論文がインターネット公表されることになった。しかし、インターネット公表の状況は完全とはいえない。個人情報保護、多重公表を禁止する学術ジャーナルへの掲載等、「やむを得ない事由」がある博士論文が多いためと推測される。博士論文全文をインターネット公表できない場合には、その要約を公表することとされているが、今回の調査の過程で、全文も要約も公表していない学位授与機関が確認された。博士論文は、高い学術的価値を有する貴重な文献資料である。教育研究成果の電子化及びOAの推進という、学位規則改正の趣旨に則り、インターネット公表の徹底が望まれる。

一方で、NDLは資料を広範に収集し、後世に伝えるという使命を持っている。しかし、電子形態の博士論文については、機関リポジトリにおけるメタデータの誤りのため、本来収集すべきものが収集されず、収集すべきでないものが収集されてしまう事態も生じている。機関リポジトリ推進委員会は「博士論文登録時に間違いやすい3項目」を作成し、注意を喚起している⁽¹¹⁾。NDLが博士論文を網羅的に収集し、後世に永く保存するために、学位授与機関におかれては、正確なメタデータ処理に協力していただくと幸いである。

- (1) NIIの学術機関リポジトリデータベース(IRDB)にメタデータを提供していること、NIIが策定したメタデータ・フォーマット“junii2”の改訂版(バージョン3.0以降)に対応していること、の2つの条件がある。
- (2) 自動収集の仕組みは次の通りである。(1) IRDBが学位授与機関の機関リポジトリから博士論文のメタデータをOAI-PMHにより収集する。(2) NDLがIRDBから博士論文のメタデータをOAI-PMHにより収集する。(3) NDLが博士論文のメタデータ内に記述されたURLに基づき博士論文の電子ファイルを収集する。
- (3) “学位規則の一部を改正する省令の施行等について(通知)(24文科高第937号 平成25年3月11日)”. 文部科学省. http://www.mext.go.jp/a_menu/koutou/daigakuin/detail/1331796.htm, (参照 2017-03-13).
- (4) 文部科学省高等教育局大学振興課.“平成25年度博士・修士・専門職学位の学位授与状況”. 文部科学省. http://www.mext.go.jp/component/a_menu/education/detail/_icsFiles/afieldfile/2017/01/26/1299723_10.pdf, (参照 2017-03-13).
- (5) NDLでは学位授与機関に対して、学位規則第12条の規定により文部科学大臣に提出する学位授与報告書の写しを、NDLにも電子メールで送付するよう依頼している。“国内博士論文の収集”. 国立国会図書館. <http://ndl.go.jp/aboutus/hakuron/index.html#chap4>, (参照 2017-03-13).
- (6) junii2の詳細は次のウェブページを参照。“メタデータ・フォーマット junii2”. 国立情報学研究所. <https://www.nii.ac.jp/irp/archive/system/junii2.html>, (参照 2017-03-13).
- (7) 著者版フラグとは、登録されている博士論文のバージョンを示すメタデータ項目である。博士論文の全文ファイルを含む場合は「ETD」と記入し、要約や要旨などの場合は「none」と記入する。“junii2ガイドラインバージョン3.1”. 国立情報学研究所. <https://www.nii.ac.jp/irp/archive/system/pdf/junii2guide-ver3.1.pdf>, (参照 2017-03-13).
- (8) “国会図書館へ提出するメタデータのフォーマットチェック”. 国立情報学研究所. http://www.nii.ac.jp/irp/archive/system/irdb_harvest.html#8, (参照 2017-03-13).
- (9) “国立国会図書館デジタルコレクション”. 国立国会図書館. <http://dl.ndl.go.jp/>, (参照 2017-03-13).
- (10) 学位規則第9条第2項の規定による。“学位規則(文科省令第23号 平成28年4月1日)”. 電子政府の総合窓口 e-Gov. <http://law.e-gov.go.jp/htmldata/S28/S28F03501000009.html>, (参照 2017-03-13).
- (11) “博士論文登録時に間違いやすい3項目”. 機関リポジトリ推進委員会. https://ir-suishin.repo.nii.ac.jp/?action=common_download_main&upload_id=879, (参照 2017-03-13).

[受理：2017-05-12]

Watanabe Jun
The Current Status and Problems Concerning the Acquisition of Doctoral Dissertations by the National Diet Library