

CA2082

国立国会図書館による 博士論文収集の現況と課題：2025年調査

にしむらよしき
西村佳樹*
いとうひびき
伊藤響*
しもむらしゅう
下村 秋†

1. はじめに

国立国会図書館（NDL）は、学術研究成果の公開・利用の促進に資するため、博士の学位の授与に係る論文（以下「博士論文」）を重要なコレクションと位置付け、網羅的な収集及び保存を行っている。

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（E1418参照）、博士論文をインターネットの利用により公表することとされた後は、主として電子形態の博士論文の収集を行ってきた。これにより各学位授与機関による分散的な保存・提供を補完し、博士論文の国レベルの長期的かつ網羅的な保存機能を担っている。

この学位規則改正及び博士論文の収集状況等に関する2017年の調査（以下「前回調査」；CA1900参照）から相当期間が経過したため、本稿では、NDLにおける収集状況を改めて報告し、併せて博士論文の収集に関する課題について述べたい。

2. NDLにおける博士論文の収集

NDLにおける博士論文の収集は、①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）⁽¹⁾と連携した自動収集、②学位授与機関からの送信（以下「送信」）、③冊子形態の送付、という三つの方法で行っており（CA1900参照）、基本的にはいずれか一つの方法でのみ収集することとなる。

しかし、このうち二つの方法で同一の博士論文をいわば重複して受領する場合がある。例えば、個人情報の保護、多重公表を禁止する学術ジャーナルへの掲載等の「やむを得ない事由」がある場合は、学位授与機関の承認を受けて、論文全文に代えてその内容を要約したものを公表することができ（学位規則第9条第2項）、そうした論文の全文は②の送信により収集されるが、その後、事由が解消して機関リポジトリに論文の全文が掲載されると、①の自動収集により同一のものが収集されることになる。現状、両者を同定して重複除外する仕組みがないため、同一の博士論文を2点とも保存することになる。そこで今回は、このような

重複を排除するため、各学位授与機関の報告番号等を人力で確認して収集件数等の調査を行った。

2.1 収集状況

文部科学省が公表している学位授与状況（令和3年度）に基づき、令和3年度の学位授与分を標本として、NDLにおける博士論文収集率を推計した。まず、文部科学省によれば、令和3年度の博士の学位授与数は1万5,767件⁽²⁾である。これに対して、NDLにおける博士論文の収集実績は表1のとおりである。

表1 博士論文の収集実績（令和3年度学位授与者）

収集方法	収集件数
自動収集	8,871
電子形態の送信	3,821
冊子形態の送付	844
合計	13,536

※異なる方法で収集した同一の博士論文は、「自動収集」により収集されたものがあれば「自動収集」として計上し、「電子形態の送信」と「冊子形態の送付」で重複した場合は、「電子形態の送信」で収集したものととして計上した。

したがって、博士の学位授与数1万5,767件に対し、NDLの収集は1万3,536件であり、NDLの収集率は85.9%、未収集率は14.1%と推計される。

2.2 未収集の原因

未収集の原因を調査するため、令和3年度分を対象に、前回調査と同じく「収集率が20%未満かつ未収集の論文が20件以上ある大学」を標本にすることとし、この条件に合致する15大学878件の博士論文を抽出した。なお、以下では主に電子形態の博士論文について言及する。

まず、博士論文の全文が機関リポジトリと大学のウェブサイトのいずれにおいても公表されておらず、送信による納入もされていないものが15大学767件あった。本文以外の公表状況は、表2のとおりである。

表2 機関リポジトリ又は大学のウェブサイトで博士論文全文が公表されていないものの公表状況ごとの件数

状況	大学数	論文数
要旨・審査の要旨・要約のみ公表	1	4
要旨・審査の要旨のみ公表	9	269
要約のみ公表	1	1
博士論文のアイテムあり（要旨等の公表なし）	1	71
博士論文のアイテムなし（機関リポジトリの検索でヒットしない）	6	361

* 国立国会図書館関西館電子図書館課

† 国立国会図書館収集書誌部収集・書誌調整課（本稿執筆時は関西館電子図書館課）

大学のウェブサイト上で要旨・審査の要旨のみ公表（機関リポジトリなし）	1	31
大学のウェブサイト上の公表なし（機関リポジトリなし）	1	30
合計	20	767

※「要約」は、課題決定、方法論、実験・解析、結論・考察など、論文全体が分かる形で内容を要約されたものを指しており、内容の要旨（アブストラクト）とは異なる。
 ※「アイテム」は、コンテンツファイル（全文ファイル等）及びメタデータから構成される情報の1単位を指す。
 ※1大学で複数の状況を有する場合があるため、大学数の合計は15と一致しない。

多くは、機関リポジトリ上で要旨のみが公開されているか、博士論文のアイテム自体がない場合であった。いずれの場合も博士論文全文が公開されていないため、学位規則で定める「やむを得ない事由」に該当するかは不明である。

次に、機関リポジトリに博士論文全文が公表されていたが、メタデータがデータ連携⁽³⁾のルールと異なった記述となっていたためハーベストエラーとなり収集できなかったものが、5大学111件あった。メタデータ項目別の内訳は表3のとおりである。

表3 メタデータがデータ連携のルールと異なった記述となっていたため、ハーベストエラーになった件数（メタデータ項目別）

メタデータ項目名	大学数	論文数
5 アクセス権 (dcterms:accessRights)	4	48
14 資源タイプ (dc:type)	2	19
35.1 本文 URL (jpcoar:file/jpcoar:URI)	5	109
合計	11	176

※1件の論文において複数項目に該当する場合があるため、各列の合計は5大学111件と一致しない。

なお、前回調査時は、標準的なメタデータスキーマとして、NIIのIRDBを介して広く流通していたjunii2が主流だったが、2017年10月にJPCOARスキーマver1.0⁽⁴⁾が、2022年12月にはJPCOARスキーマver2.0が公開され、現在はこちらが主流となっている⁽⁵⁾。

2.3 収集対象外データの連携

NDLが機関リポジトリから自動収集した博士論文のうち、本来収集対象ではないファイルがデータ連携されたケースについて、調査時点で直近となる令和5年度の間に確認した分⁽⁶⁾を対象に調査を行った。結果、

自動収集されたデータ1万6,166件のうち、収集対象外のデータは6,076件であり、37.6%に相当した（なお、収集対象外のデータは保存・公開しておらず、表1にも計上していない）。収集対象外となる主な理由は次のとおりである。

- ・公開予定のステータスであるために、NDL側で全文ファイルをダウンロードできず、メタデータや要旨のみが収集された。
- ・機関リポジトリのシステムリニューアル、アイテムの削除・再登録の操作等により、連携されるメタデータの更新のキーに用いる識別子（OAI-PMHのidentifier）が変更され、収集済みの博士論文のメタデータが新規アイテムとして判定され、再収集される。
- ・全文が公開されていないが、junii2又はJPCOARスキーマの条件を満たしているためメタデータ等のみが収集された。

前回調査における割合は8.7%であったため、今回の調査では割合が大幅に増加したことになる。JAIRO Cloudのバージョンアップや、データ収集に関与するNDLのシステムリニューアルによるデータ連携不備の影響、各機関における大量のデータ作成・更新が原因と考えられる。また、自動収集後の確認作業において、次のような問題が発見されることがある。

- ・全文PDFに印刷不可の設定が付与されている⁽⁷⁾。
- ・全文PDFにデータ破損、被覆等がある。
- ・別の著者の博士論文が誤って登録されている。

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又は論文ファイル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

改正後の学位規則では、大学の教育研究の成果である博士論文等の質を相互に保証し合う仕組みとして、博士論文等を相互に参照できるよう公表することを規定している。一方で、前述のとおり、「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（学位規則第9条第2項）。そこで「2.1 収集状況」及び「2.2 未収集の原因」で調査した結果を基に、博士論文の全文公表状況を推計した。

NDLが収集した令和3年度分の博士論文を次のとおり類型化し、「公表」又は「非公表」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
 - ・「部分参照 (URI)」入力あり→全部について公表有無調査を実施。
 - ・「部分参照 (URI)」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→標本調査を実施し、公表率を直接推計。

(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は、メタデータ項目の「部分参照 (URI)」に、論文ファイルを掲載しているページの URL を入力するように依頼している。電子形態の送信により収集した博士論文は 3,821 件であり、このうち「部分参照 (URI)」の入力があるものは 40 件であった。この 40 件を全数調査したところ、全件ともインターネットで全文が公表されていたが、うち 26 件は自動収集でも収集しており、残りの 14 件は大学ウェブサイト上で公表されていた。よって、14 件をこの類型における「公表」と判定した。

(4) について、「2.2 未収集の原因」の調査結果から推計した。未収集として調査した 878 件のうち、機関リポジトリ又は大学のウェブサイトで博士論文全文が公表されていない 767 件 (表 2) は「未公表」、メタデータ記述が原因でハーベストエラーとなった 111 件は全文が機関リポジトリで公表されていたため「公表」と判定した。そのため、この類型における公表率は、調査対象 878 件に対する 111 件であり、12.6%と推計した。

以上の結果は表 4 のとおりである。令和 3 年度分を対象とした調査では、博士論文全文のインターネット公表率は 58.1%程度と推測された。

表 4 博士論文全文のインターネット公表率の推計 (令和 3 年度学位授与者分)

	件数	公表件数	公表率
自動収集	8,871	8,871	100%
電子形態の送信	3,821	14	0.4%
冊子形態の送付	844	0	0%
NDL 未収集	2,231	(281)	12.6%
全体	15,767	9,166	58.1%

※公表率は小数点 2 位以下を四捨五入。

括弧内の数字は推計した公表率から逆算で算出。

大学院における教育研究成果の電子化及びオープンアクセスの推進の観点から、博士論文は可能な限りイ

ンターネットで公表されることが求められているところであるが、表 4 の結果からは、4 割強のものが学位規則第 9 条第 2 項の適用対象となり全文が公表されていないと推測される。また、送信により収集した博士論文のうち一定数⁽⁸⁾については学位授与機関からの申出を受けて館内閲覧や複写に関する利用制限措置を行っている。博士論文の利用に当たっては、これらの状況にも留意されたい。

4. 課題

NDL による博士論文の収集に関する課題として、次の点が挙げられる。

・自動収集におけるデータ連携

博士論文の自動収集は、国立国会図書館サーチ (以下「NDL サーチ」) が IRDB から収集したメタデータのうち、自動収集の条件に合致する博士論文のメタデータを国立国会図書館デジタルコレクションに取り込むことにより行っている⁽⁹⁾。この際に、「2.3 収集対象外データの連携」で述べたとおり、メタデータ記述がデータ連携のルールに合致していたとしても、全文ファイルのステータスが公開予定であるため取得できず要旨のみが収集されたり、アイテムの削除・新規作成により収集済の博士論文が新規扱いで収集されたりするなど、データ連携上の課題が多い。NDL サーチの連携条件の変更等により収集対象外データの割合を減らせることが望ましいと考えているが、詳細は検討中である。

・重複収集

「2. NDL における博士論文の収集」で述べたとおり、同一の博士論文を重複して収集・保存するケースがある。手作業による重複確認なども考えられるが、自動収集はメタデータが更新されるたびに発生するほか、各機関リポジトリにおける運用変更等による大量更新も起こるため、確認が必要となる件数が多く、それら全てを人力で確認し続けることは困難と思われる。当館が収集する際に機械的に同定することも検討しているが、同定を行うための情報が十分でない⁽¹⁰⁾ため、学位授与機関における DOI の付与等、メタデータの整備・充実にも期待したい。

・収集の漏れ

「2.1 収集状況」に記載のとおり、博士論文全体の約 14.1%が NDL 未収集と推計される。また、表 4 のとおり、博士論文全文のインターネット公表率は 6 割未満となっており、未公表のものは学位授与機関からの送信により収集を進める必要がある。今回の調査によ

り未収集が確認できた大学に対しては、納入依頼等を行うことを検討している。

5. おわりに

本稿では、NDLの博士論文収集の状況を紹介するとともに課題について述べた。

各学位授与機関において機関リポジトリの維持管理及びNDLへの送信を行うことは相当な負担になっているものと推察されるが、学術成果の長期的な保存及び利用提供に資するため、今後も当館事業へのご協力をお願いしたい。

- (1) 日本国内の学術機関リポジトリに登録されたコンテンツのメタデータを収集し、提供するデータベース・サービス。
“IRDBとは”. 学術機関リポジトリデータベースサポート.
<https://support.irdb.nii.ac.jp/ja>. (参照 2025-04-22).
- (2) 1万5,767件には、大学改革支援・学位授与機構が行った学位授与の数は含まれていない。
“令和3年度博士・修士・専門職学位の学位授与状況”. 文部科学省.
https://www.mext.go.jp/a_menu/koutou/daigakuin/detail/mext_0251203.html. (参照 2025-04-22).
- (3) “データ連携 - 国立国会図書館”. 学術機関リポジトリデータベースサポート.
https://support.irdb.nii.ac.jp/ja/harvest/jpcoar/dataprovide_ndl. (参照 2025-04-22).
- (4) 大園隼彦ほか. JPCOARスキーマの策定:日本の学術成果の円滑な国際的流通を目指して. 情報管理. 2018, 60(10), p. 719-729.
<https://doi.org/10.1241/johokanri.60.719>. (参照 2025-04-22).
- (5) 2024年度オープンアクセスリポジトリ推進協会(Japan Consortium for Open Access Repository:JPCOAR) 総会資料によると、国公私立大学のうち、約95%の機関が「JAIRO Cloud」を利用している。JAIRO Cloudは、JPCOAR及びNIIが共同運営しているクラウド型の機関リポジトリサービスであり、その基盤となるソフトウェア「WEKO」は2020年から2023年にかけて、WEKO2からWEKO3へバージョンアップが行われた。WEKO3は、標準メタデータスキーマとして、JPCOARスキーマを採用している。
“2024年度JPCOAR総会”. JPCOAR.
<https://jpcoar.repo.nii.ac.jp/records/2000581>. (参照2025-04-22).
- (6) 各学位授与機関で博士論文の公開作業時期が異なるため、令和5年度以前に学位授与された分(複数年度)の博士論文が含まれる。
- (7) 長期的な保存及び利用に適した形式とするため、印刷制限等を行わない形式で送付等をしていただくことになっている。
- (8) 本文中で具体的なデータは示していないが、令和3年度学位授与者(送信分)について、令和7年3月時点で措置中の実績値は約25%である。
- (9) 植山廣紀. “「国内博士論文の収集」の舞台裏！～国立国会図書館インタビュー”. JPCOARウェブマガジン. 2024-10-29.
<https://magazine.jpcoar.org/news/17822e4f-c842-49c6-acd2-69e44b5b4e95>. (参照 2025-04-22).
- (10) 日英記載等の関係で、自動収集と電子形態の送信の場合で、タイトル・著者名・学位授与機関名等の表記揺れがある。

[受理: 2025-05-23]

Nishimura Yoshiki, Ito Hibiki, Shimomura Shu
The Current Status and Problems Concerning the
Acquisition of Doctoral Dissertations by the National
Diet Library: Reporting in 2025