

機関リポジトリにおける博士論文のデータ入力状況調査

平成 28 年 3 月

機関リポジトリ推進委員会
コンテンツワーキンググループ

概要

各機関において機関リポジトリに登録した博士論文が、**JAIRO** や国立国会図書館デジタルコレクション、**CiNii Dissertations** に適切に反映されていない場合がある。そこで、各機関リポジトリのデータと **IRDB**、**JAIRO** のフォーマットの対応状況を調査することで、その原因を探り、改善のためにリポジトリ担当者が取りうる対処方法を検討した。また **junii2** の改訂に向け、追加項目の有無についても検討を行った。

1. はじめに

学位規則の改正により、2013 年 4 月に博士論文のインターネット公表が義務化されてから 2 年を経過している。昨年度の機関リポジトリ推進委員会コンテンツワーキンググループの調査でも、各機関で博士論文が順次公開されていることが明らかとなっている。しかし、各機関が機関リポジトリに登録したデータが、**JAIRO** や国立国会図書館デジタルコレクション、**CiNii Dissertations** への反映されていないケースが見受けられる。入力方法の違い等に起因すると考えられるが、何が原因であるかについてこれまで調査されておらず、全体像は不明である。

本調査では、各機関の機関リポジトリに入力したデータが他のシステムに連携されない原因を探り、その改善のためにリポジトリ担当者が取りうる対処方法を検討するため、各機関の機関リポジトリと **IRDB** (**Institutional Repositories DataBase** : 学術機関リポジトリデータベース)、**JAIRO** のデータベースを対象に、各データの対応状況調査を行った。また、**junii2** の改訂に向けて、博士論文に必要な項目追加の有無についても調査を実施した。

2. 対象データ

本調査においては 2001 年から 2015 年に登録された博士論文を対象として、調査した。対象を特定するために、次の手順でデータを抽出した。

最初に、国立情報学研究所担当者により、平成 28 年 1 月 4 日に抽出された、以下のデータの提供を受けた。

- ・対象データベース : **IRDB**
- ・抽出条件 : **NIItpe: Thesis or Dissertations AND [(dateofissued: NULL AND dateofgranted: NULL) OR dateofissued: 2001-2015 OR dateofgranted:**

2001-2015]

- ・抽出結果件数：165,605 件

次に、IRDB データのうち、以下の a～e の条件のいずれかに該当するものを博士論文に該当するものとしてデータを抽出した。

- ・ degreename に記載があるもの
 - a) degreename = 「博士」または機関リポジトリのデータを確認して博士論文であるとわかったもの
- ・ degreename が空欄のもの
 - b) grantid = 「甲」、「乙」、「A」、「B」のいずれかを含むもの
 - c) naid = 「博士」のもの
 - d) jtitle = 「博士論文」または「博論」を含むもの
 - e) いずれかの項目に「ETD」を含むもの

その結果、対象データ件数は 85,186 件となった。

3. 調査方法

本調査では、別紙 1 の各機関について、1 年度あたり 5 件程度のデータをサンプル抽出し、各機関の機関リポジトリの Web サイトと IRDB データ、JAIRO の Web サイトの項目を比較し、データの反映状況を確認した。併せて、junii2 改訂に向けて、博士論文に関する項目で、各機関がローカルで用いている項目等の中に追加を検討すべきものがないかについても調査を行った。対象機関の選出の方法は以下の通りである。ソフトウェア毎の傾向をみるため、各機関が用いているリポジトリのソフトウェアを IRDB で調査し、設置機関の別や規模等も考慮して、ソフトウェア毎に調査対象機関を選出した。

また、インターネット公表義務化前後のフォーマット変更等も考慮する必要があることから、本調査では 2001 年、2009 年、2013 年を調査対象年とし、該当年がない機関については、これに近い年度を対象とし、IRDB データで年度が不明なものは「不明」とした。結果、調査対象機関は 47 機関で、対象年度は別紙 1 に示すとおりとなった（ただし、1 大学については詳細情報にアクセスできなかったため、調査開始後分析対象外とした）。

4. 結果・考察

4.1. IRDB 及び JAIRO へのデータ反映状況

junii2 (バージョン 3.1) の博士論文に関する入力必須項目（「あれば必須」を含む）の IRDB 及び JAIRO へのデータ登録状況は、表 1 のとおりであった¹。「タイトル」、「作成者」、「URI」、「本文フルテキストへのリンク」、「学位名」、「学位授与機関」は全機関で入力され

¹ 詳細については別紙 2 の「IRDB 対応表」シートを参照のこと。

ていたが、「国立情報学研究所 メタデータ主題語彙集（資源タイプ）」、「著者版フラグ」、「学位授与番号」についてはきちんと登録されていない機関があった。

データの反映状況についても、「国立情報学研究所 メタデータ主題語彙集（資源タイプ）」、「著者版フラグ」、「学位授与番号」については、IRDB や JAIRO に上手く反映されていない例が見られた。特に「学位授与番号」については、IRDB や JAIRO にデータが上手く反映されていない機関が少なくとも 7 機関²と突出していた。その内訳としては、機関リポジトリに「学位記番号」の登録があるが「報告番号」が欠落しているため IRDB の「grantid」に上手く反映されていないものや、「学位授与番号」が登録されているにもかかわらず IRDB の「grantid」にデータが反映されていないもの、または「報告番号」だけ登録されているものなどであった。ソフトウェアによる差異は見られないが、2011 年以降のデータしかない機関を中心に見られた。このことから、特に近年機関リポジトリを立ち上げた機関において、「学位授与番号」の正しいフォーマットが把握できておらず、入力もれや誤りが生じやすい状況にあると考えられる。

このほか、「学位授与機関」等を機関リポジトリ上では表示させず、IRDB や JAIRO にのみ内容を反映させている機関も見られた。これは各機関における必要性の判断により、非表示にしているためと思われる。

表 1 IRDB や JAIRO へのデータ登録状況³ (N=46)

博士論文に関する junii2 項目名	必須	機関リポ ジトリ登 録機関数	IRDB デ ータ反映 機関数	JAIRO デ ータ反映 機関数
タイトル	必須	46	46	46
作成者	あれば必須	46		46
国立情報学研究所 メタデータ 主題語彙集（資源タイプ）	必須	42		43
URI	必須	30	46	
本文フルテキストへのリンク	あれば必須	46	46	46
著者版フラグ	あれば必須	39	44	44
学位授与番号	あれば必須	43	45	45
学位授与年月日	あれば必須	45	45	45
学位名	あれば必須	44	46	46
学位授与機関	あれば必須	43	46	46

また、「必須」もしくは、「あれば必須」項目以外では、「言語」に関する記述が機関リポ

² 詳細については別紙 2 の「IRDB 反映状況（一部）」を参照のこと。

³ ここでは 1 年度でも登録があればカウントした。

ジトリにあるにも関わらず、IRDB の「title 属性」に反映されていないケースが多数見られた。

4.2. junii2 への追加項目調査

各機関でローカル項目に用いている項目のうち、ローカル項目のみで用いられているものを表 2 にまとめた。ここでは、各種日付やローカルで使用している ID、junii2 の項目を別途記載しているもの（例えば「学位授与番号」とは別に「報告番号」や「学位記番号」を記載している等）、junii2 の項目に詳細を記載すれば足りる項目は除外している。「論文調査員（審査員）」が 7 機関と最も多く、次いで「ファイルサイズ」が 4 機関、「請求記号」と「閲覧数」が 3 機関となっていた。ただし、「ファイルサイズ」と「閲覧数」については、ほとんどが E-Cats Library を使用している館であったため、E-Cats Library の仕様が影響している可能性がある。「論文調査員（審査員）」については、junii2 の「寄与者」を利用することは可能と思われるが、役割を明示して記載する必要があるため別途項目立てしているものと思われる。近年論文の質が厳しく問われてきていることもあり、「論文調査員（審査員）」を junii2 改定時に追加対象として検討することを提案したい。

表 2 ローカル項目使用状況

各機関のローカル項目	機関数
論文調査員（審査員）	7
査読有無	1
公開状態（学外・学内の別）	1
学位授与の要件	1
請求記号	3
ファイルサイズ	4
閲覧数	3
NDL ID	1

5. 総括

機関リポジトリに登録した博士論文のデータが正しくハーベストされ、他システムに反映されるようにするため、IRDB 及び JAIRO へのデータ反映状況調査を行った。junii2 の「必須」と「あれば必須項目」においては、特に近年機関リポジトリを導入した館で「学位授与番号」が適切に反映されない状況が見られた。「学位授与番号」は、「科研費機関番号 (5 桁) + [甲 | 乙 | *] + 第* + 報告番号 + 号」の形式で記述する」という書式で入力することになっているが、「学位記番号」「報告番号」等関連する番号が複数あることも影響し、こ

の点についての登録担当者の理解不足が原因と思われる。対処方法としては、該当機関への連絡や、機関リポジトリ新任担当者研修等で強調して周知していくことが考えられる。

また、junii2改定に向けて、博士論文に関する項目で追加を検討すべき項目の有無についても調査を行った。その結果「論文調査員（審査員）」がローカル項目で最も使われていたことが判明した。論文の質を担保するという意味も含めて、「論文調査員（審査員）」を追加項目として検討対象とすることを提案したい。

以上