情報知識学会 第20回年次大会
研究発表 2012.5.20

日本の学協会誌掲載論文の機関リポジトリ収録状況

清水真理, ◯佐藤翔, 逸村裕

1

それでは「日本の学協会誌掲載論文の機関リポジトリ収録状況」と題して発表します。
筑波大学図書館情報メディア研究科の佐藤翔です。

目次

1.はじめに：研究背景と目的

2.方法・対象

3.分析結果

4.考察と今後の課題

2

本発表の概要はこちらの通りです。

目次

1.はじめに：研究背景と目的

2.方法・対象

3.分析結果

4.考察と今後の課題

3

はじめに、研究の背景と目的ですが。

本研究の目的

- 日本の学協会が発行する雑誌に掲載された論文の、機関リポジトリ収録状況を調査
- 全体及び著作権ポリシー、学術分野との関係も明らかに

4

本研究の目的は、日本の学協会が発行する雑誌掲載論文のうち、どれだけが機関リポジトリに収録されているかを明らかにすることです。
全体の傾向だけではなく、学協会の著作権ポリシーや、分野との関係も分析します。

機関リポジトリとは？

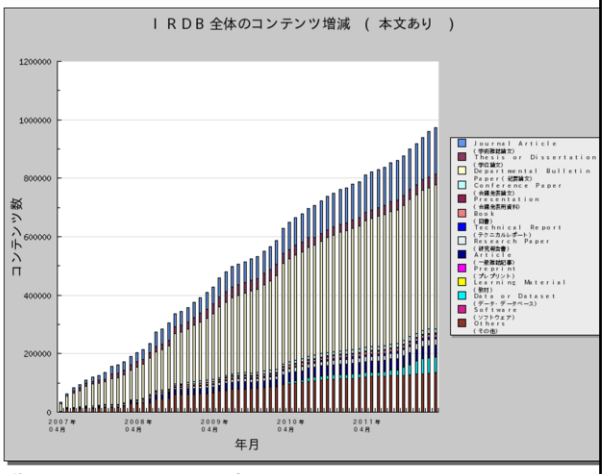
- 学術機関が、機関で生産されたコンテンツを、収集・管理・発信するシステム（サービス）
- オープンアクセス運動の一翼を担うものとして2001年頃にあらわれる
- 2005年頃から国内でも普及

5

機関リポジトリとは、学術機関が、構成員のコンテンツを収集・管理・発信するシステムです。
学術論文の自由な流通を目指すオープンアクセス運動の一翼を担うものとして、2001年頃に現れた取り組みで、日本では2005年頃から普及してきました。

機関リポジトリの整備・拡充

- 機関リポジトリ数
  - 世界全体：1,788[1]
  - 国内：169 [2]
- 収録文献数
  - 国内:約97万[3]
  - １年で約19万件増



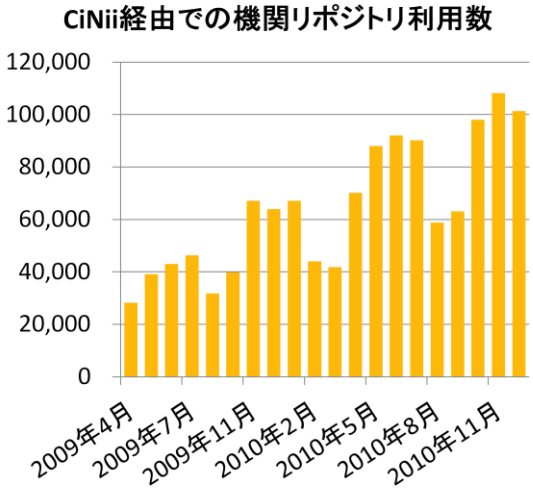
出典：IRDBコンテンツ分析システム(http://irdb.nii.ac.jp/)[3]

6

2012年4月現在、世界で1,788、日本では169機関がリポジトリを持っています。
日本のリポジトリ収録文献は97万件を超え、この1年で約19万件増えています。

機関リポジトリ提供論文の利用

- CiNii上での利用[4]
  - 月間10万回以上
  - NII-ELSを補う
- 非研究者からも盛んに利用される[5]
  - Wikipedia
  - Q&Aサイト
  - 個人ブログ



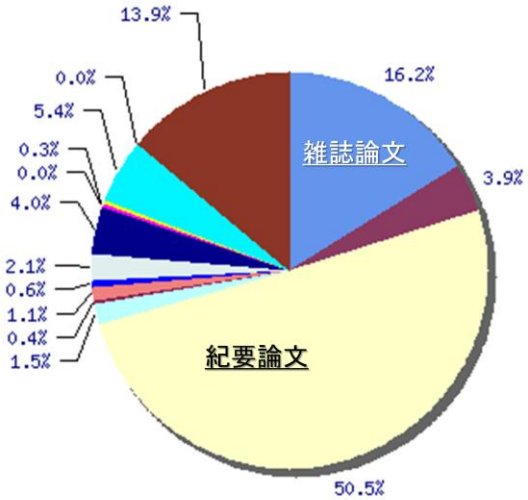
出典：佐藤ら（2012）[4]を加工

7

これらの論文は日本の学術情報流通に貢献しています。
例えば先の日本図書館情報学会でも発表しましたが、機関リポジトリ収録論文は2010年末時点で月間10万回以上、CiNii経由で利用されており、NII-ELSの足りない部分を補っていました。
また、Wikipedia、Q&Aサイト、ブログなど、研究者以外の人びとの活動の中でもリポジトリ収録論文が活用されていることがわかっています。

収録コンテンツの内訳

- 大半は紀要論文
  - 紀要論文：50.5%
  - 雑誌論文：16.2%
- 紀要は著作権処理が容易
- 雑誌論文は処理に手間がかかる[6]



出典：IRDBコンテンツ分析システム(http://irdb.nii.ac.jp/)[3]

8

このように収録コンテンツが増加し、よく使われている一方で、その中身には偏りがあることも知られています。
これは2012年4月現在の日本の機関リポジトリ収録コンテンツの内訳を見たものですが、実に過半数は大学・研究機関の発行する紀要掲載論文です。
オープンアクセス運動においては雑誌論文の収録が重要と言われていますが、日本の機関リポジトリでは雑誌論文はコンテンツの約16%にとどまります。
この理由として、紀要は発行機関が学内にあるため著作権処理が容易な一方、雑誌論文は処理に手間がかかることが指摘されています。

出典：学協会著作権ポリシーデータベース（http://scpj.tulips.tsukuba.ac.jp/）[7]

そこで著作権処理の手間を減らし、機関リポジトリへの収録を促進するために、出版社や学協会が著作権についてどのようなポリシーを持っているかを示す、データベースの作成が進められています。
その一例として、日本では日本の学協会とその発行誌を対象とする、学協会著作権ポリシーデータベース、通称SCPJがあります。
SCPJでは学協会名や雑誌名で著作権ポリシーを検索し、ある論文をリポジトリ等に収録できるかどうか調べることができます。

雑誌論文の機関リポジトリ収録

- SCPJ等によって進んでいる？
  - 効果は未検証
- そもそも・・・日本の論文の機関リポジトリ収録状況が不明
  - 英文誌掲載論文は概算あり：11.1%[8]
  - 和文誌掲載分は概算もない

10

このような取り組みによって雑誌論文の機関リポジトリへの収録は従来よりも盛んになっていると考えられますが、その効果の検証はこれまで行われていません。
それ以前に、そもそも日本の論文のうち、どれだけが機関リポジトリに収録されているのか自体、詳細な調査はありませんでした。
英文誌掲載分については11.1%と概算を示した調査もありますが、あくまで概算で実際の収録状況を見たものではなく、また和文誌については概算すらありません。

本研究の目的

- 日本の学協会が発行する雑誌に掲載された論文の、機関リポジトリ収録状況を調査
  - SCPJ収録誌を対象とする
- 全体及び著作権ポリシー、学術分野との関係も明らかに
  - 日本の機関リポジトリの現状を理解する基礎データに

11

そこで本研究ではSCPJ収録誌を対象に、日本の学協会が発行する雑誌に掲載された論文が、どれだけ機関リポジトリに収録されているかを調査しました。
この調査により、日本の機関リポジトリの現状を理解するための基礎データを提供できると考えられます。

目次

1.はじめに：研究背景と目的

2.方法・対象

3.分析結果

4.考察と今後の課題

12

続いて分析方法と対象についてです。

分析対象論文

- 対象雑誌：SCPJ収録誌中、ISSN記載のあった2,583誌
- 対象期間：2000～2009年掲載論文
- CiNii APIを用い論文データ取得
  - 一度に200件までデータ取得可能
  - 年間200本以上掲載する雑誌・・・
    新しいもの200本まで取得

13

本研究で対象とするのはSCPJに収録されていた雑誌の中で、ISSNの記載があった2,583誌に、2000～2009年の10年間に掲載された論文です。
論文データはCiNiiのAPIにより取得します。
その際、APIの制限から、1年間に200本以上論文を掲載する雑誌については、新しい順に200件まで対象としました。

分析対象論文

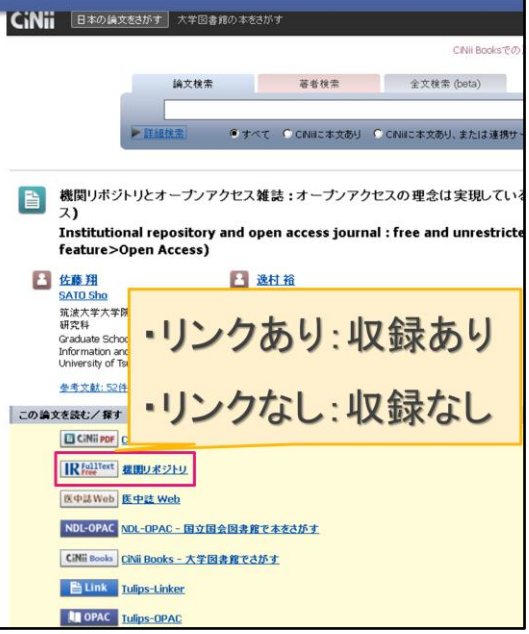
- データは2011年9月に取得
- 463誌はCiNiiに論文掲載なし：除外
- 最終的な分析対象：
  - 2,120誌
  - 1,010,822論文

14

データは2011年9月に取得しました。
データを確認したところ、463の雑誌についてはCiNiiに論文が1本も掲載されていませんでした。これらは廃刊・休刊やCiNiiに未収録の雑誌と考えられます。
この463誌は分析から除外し、最終的な分析対象は2,120の雑誌に掲載された、1,010,822本の論文となりました。

分析方法:リポジトリ収録

- 全体および著作権ポリシー、学術分野別の機関リポジトリ収録状況を分析
- リポジトリ収録状況：CiNiiからのリンクの有無で判断



これらの論文全体、および著作権ポリシーや学術分野別に、機関リポジトリへの論文収録状況を分析します。
機関リポジトリへの論文収録の有無は、CiNiiから機関リポジトリへのリンクの有無から判断しました。
今、画面右に出しているのが実際のCiNiiの画面ですが、この論文のように機関リポジトリへのリンクがあるものは「機関リポジトリ収録あり」、リンクのないものは「収録なし」と判断します。なお、リンクの有無についてもAPIで、論文データとあわせて取得しました。

著作権ポリシーについては分析の元であるSCPJデータベースのデータを用います。
SCPJでは画面のように、学協会の著作権ポリシーを緑、青等の色であらわします。
緑は査読前・査読後のどちらも機関リポジトリに収録してよい、青は査読後のみOK,黄色は査読前のみOK、白は機関リポジトリへの登録不許可、灰色は検討中や無回答です。

本研究ではこれをさらになんらかのバージョンを機関リポジトリに収録して良い「収録可」と、リポジトリへの収録を認めない、もしくはまだ決まっていない「収録不可」の２つに分け、それぞれの傾向を分析しました。

学術分野データ

- 学会名鑑の分野わけを参照
  - 記載のない学会は掲載論文の内容に基づき判断
  - 人文・社会、生命科学、理学・工学の3領域、30分野＋「複合領域」
- 詳細は予稿表1参照

18

また、学術分野については学会名鑑の分野わけを参考にしました。
学会名鑑未掲載の学会についてはその雑誌掲載論文の内容に基づき分野を判断し、人文・社会、生命科学、理学・工学の3領域、30分野＋複合領域に分けました。
詳細はお手元の予稿表１を御覧ください。

目次

1.はじめに：研究背景と目的

2.方法・対象

3.分析結果

4.考察と今後の課題

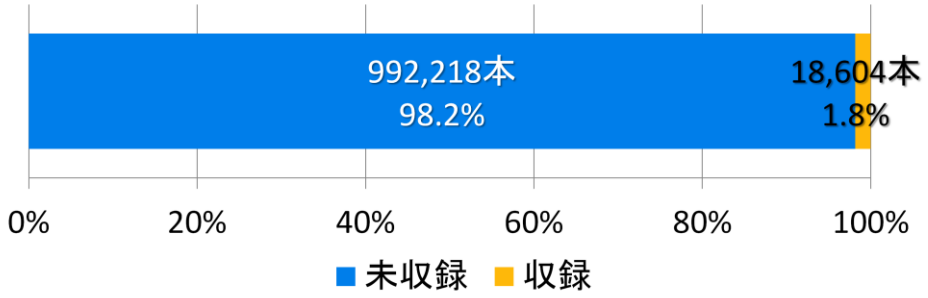
19

以上の方法に基づいて分析した結果についてですが、

(1)機関リポジトリへの論文収録

まず機関リポジトリへの論文収録状況、全体の様子を見ると

機関リポジトリ収録状況



- ほとんどの論文はリポジトリ未収録
- 収録されているのは1.8%にとどまる

21

これは分析対象の約100万件の論文中、機関リポジトリに収録されていなかったものを青、収録されていたものを黄色で示した帯グラフです。
黄色がほとんどなく、全論文のうち機関リポジトリに収録されていたのは1.8%と、ごく少数にとどまっていました。

収録数上位タイトル

1. 『物性研究』（1,681 / 1,861本収録）
   - 京都大学基礎物理学研究所が刊行
2. 『千葉医学雑誌』（1,380 / 1,442本収録）
3. 『東京慈恵会医科大学雑誌』（939 / 1,097本収録）
4. 『The KITAKANTO Medical Journal』（924 / 1,670本収録）
   - 群馬大学内の学会

22

さらにその1.8%のうち、特に機関リポジトリ収録論文数の多い雑誌上位4位がこちらです。
1位の『物性研究』は京大基礎物理研の雑誌、2位・3位は千葉大と東京慈恵大の医学雑誌、4位は、北関東医学会という、群馬大学に本拠を置き、医学系の研究科長が会長を兼ねる学会の雑誌です。

収録数上位タイトル

1. 『物性研究』（1,681 / 1,861本収録）
   - 京都大学基礎物理学研究所が刊行
2. 『千葉医学雑誌』（1,380 / 1,442本収録）
3. 『東京慈恵会医科大学雑誌』（939 /

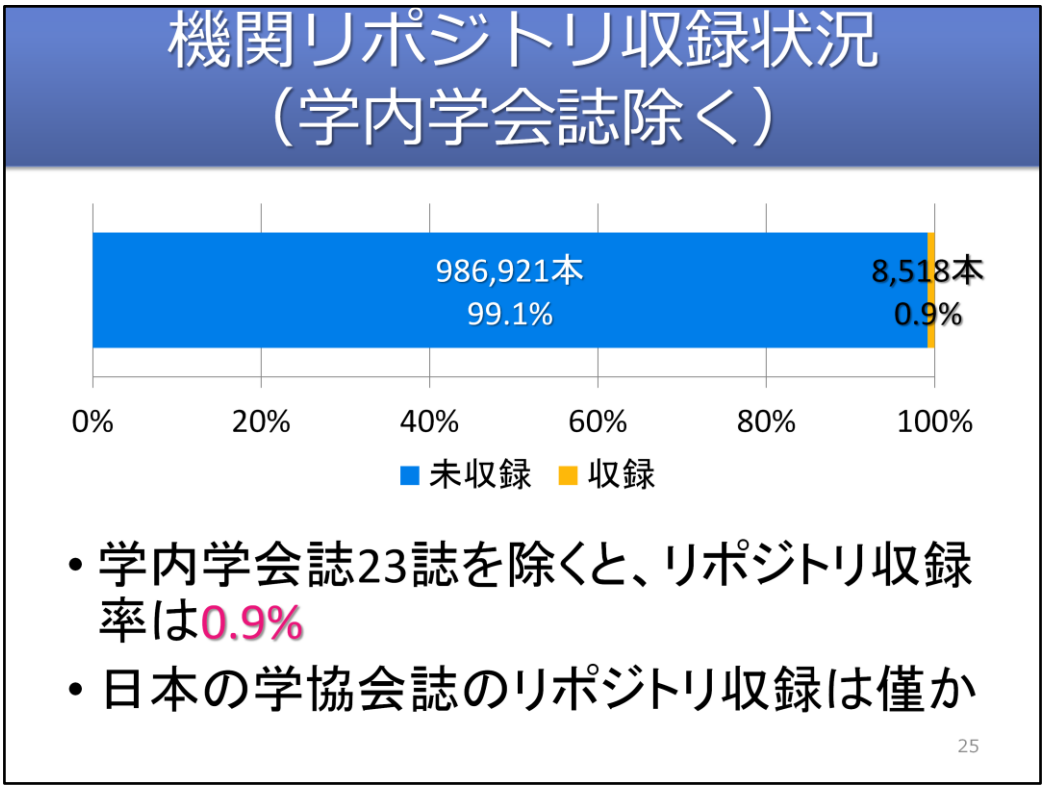
・いわゆる「学内学会」誌が上位を占める

・特定の（学会が属する大学の）リポジトリに収録

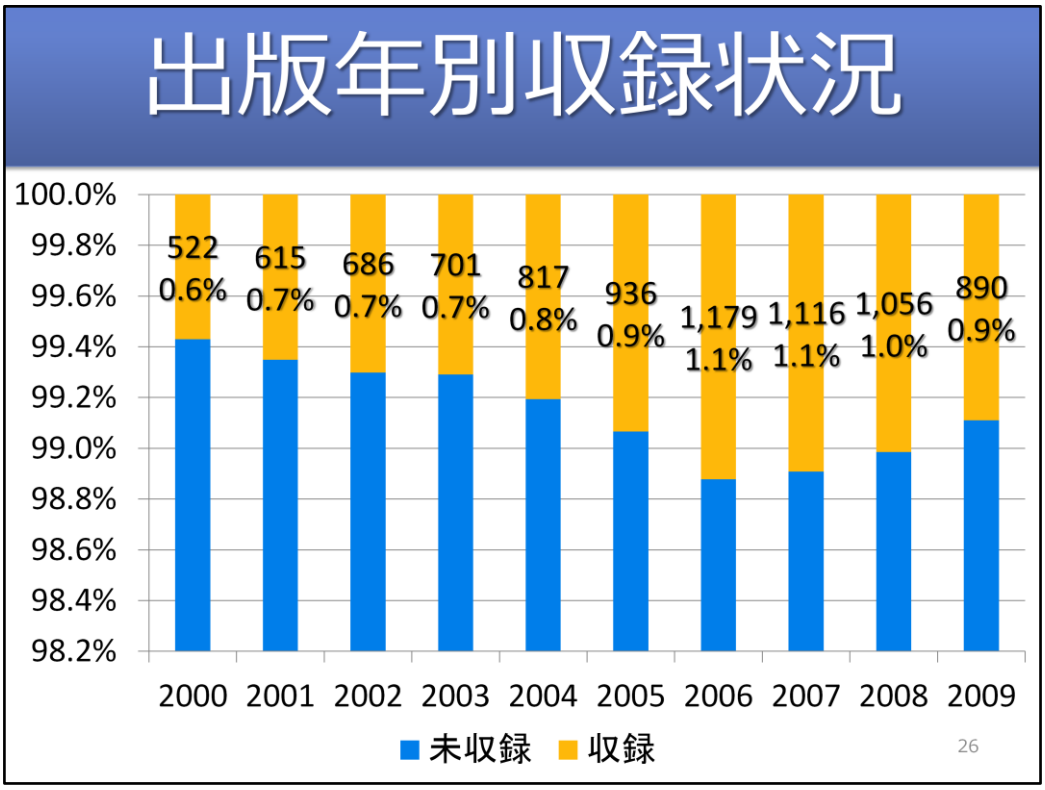
⇒・実質的に紀要と同様？

この他にも機関リポジトリに論文が多く入っているのはほとんどがいわゆる学内学会の雑誌、特定の大学所属者によって刊行されている雑誌で、論文のほとんどがその大学のリポジトリのみに登録されていました。
これは実質的に紀要と同様と考えられるものです。

実際、これは1位の『物性研究』の機関リポジトリ上のデータを見たものですが、学会誌ではなく「Departmental Bulletin Paper」、紀要論文として登録されています。このような雑誌は、学協会誌掲載論文の収録状況を分析する、という本研究の意図からは外れたものです。



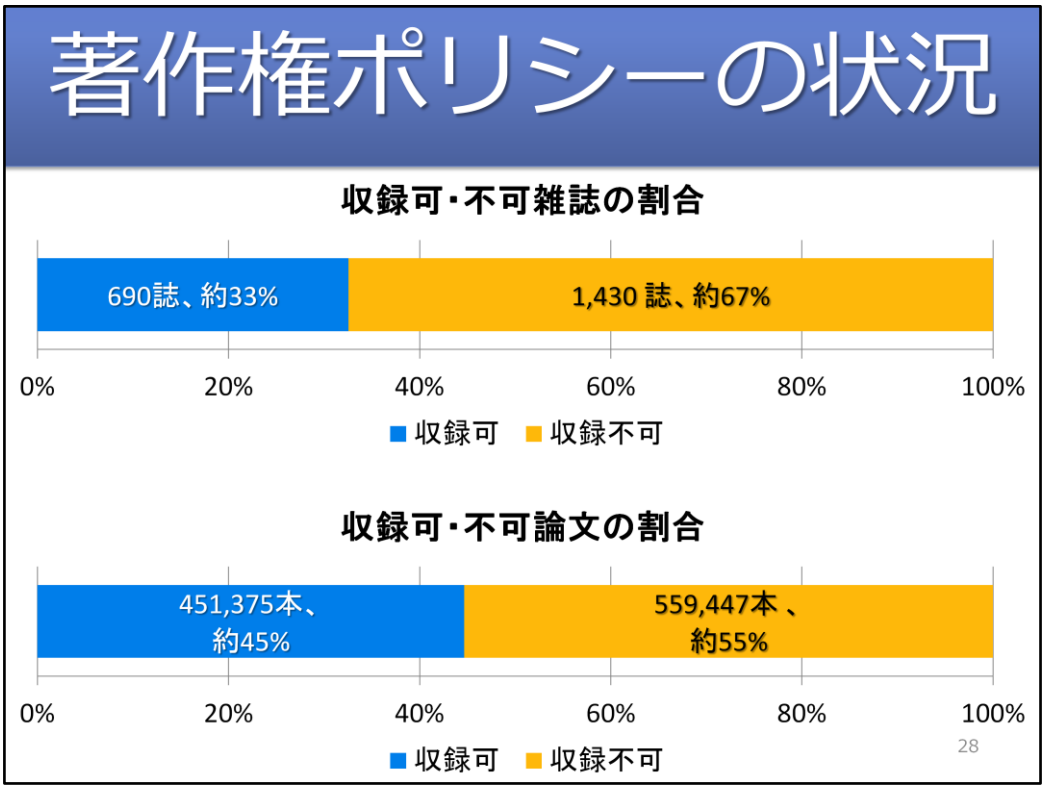
そこで特に機関リポジトリ収録論文数の多い学内学会誌23誌を除いて分析しなおしたのがこちらのグラフです。学内学会誌を除くと、機関リポジトリ収録論文は全体の0.9%と、1%を割り込むまでに減ります。
日本の学協会誌に掲載された論文のうち、機関リポジトリに収録されているものはごくわずかである、と言えるでしょう。



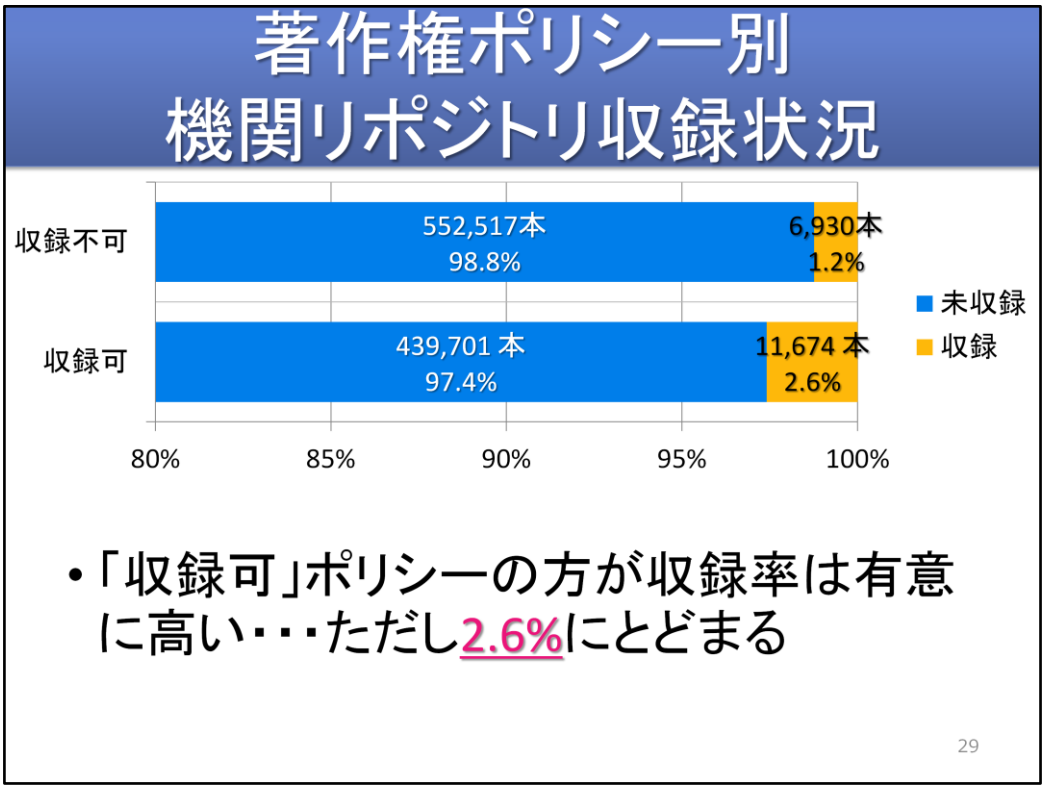
ちなみにこれは論文の出版年別に収録状況を示したものです。
2006年くらいまでは新しいものほど収録されている割合が高い傾向がありますが、そのあとは徐々に減ってきています。
これは出版後、機関リポジトリへの論文収録を一定期間猶予するよう求める、いわゆるエンバーゴを設定している学会の影響と考えられます。

(2)著作権ポリシーとの関係

次に著作権ポリシー、機関リポジトリへの収録を学会が認めているか否かと実際の収録状況の関係ですが。

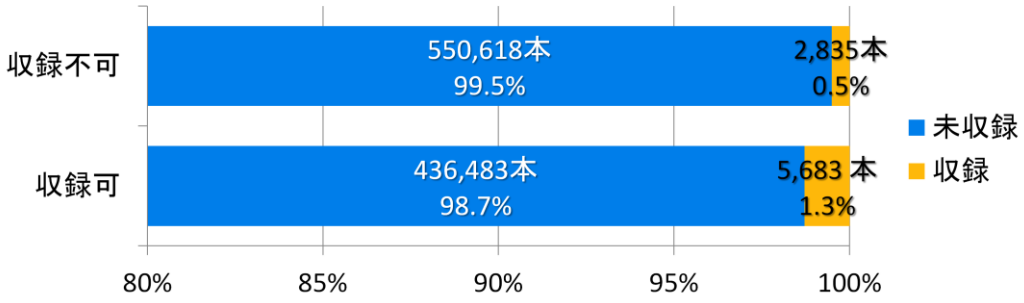


まずそもそもの著作権ポリシーの状況について、本研究の対象雑誌・論文の状況を見たものがこちらです。
雑誌単位では収録を認めているものは全体の33%にとどまっているのですが、論文単位で見ると全体の45%と半数近くは機関リポジトリへの論文収録が認められたものでした。
これは収録を許可する雑誌の中に、大規模雑誌が多いことの影響と考えられます。



次にこの著作権ポリシーごとに、機関リポジトリへの収録状況を見たものがこちらです。
上が収録が許可されていない論文、下が収録が許可された論文の収録状況ですが、当然といえば当然ながら、機関リポジトリへの収録が認められたものの方がよく収録されています。X二乗検定より、この差は有意でした。
ただし、「機関リポジトリに収録しても良い」論文でも実際の収録率は2.6%にとどまっています。

著作権ポリシー別機関リポジトリ収録状況（学内学会誌除く）



- 学内学会誌を除くと「収録可」でも機関リポジトリ収録率は1.3%
- 「収録可」論文のほとんどは未収録

30

さらに先の数字は学内学会誌を含んでいたのですが、それを除いた場合には「機関リポジトリへの収録可」の論文でも、実際に収録されているのはわずか1.3%と半減しました。
たとえ学協会がリポジトリへの収録を認めている場合であっても、ほとんどの論文はリポジトリに収録されていないことがわかりました。

(3)分野との関係

最後に、学術分野との関係を見ていきます。

領域別機関リポジトリ収録状況

領域	総論文数	収録許可論文数	収録許可率	リポジトリ収録数	リポジトリ収録率
人文・社会	147,221	49,168	33.4%	1,415	1.0%
生命科学	484,580	182,299	37.6%	3,570	0.7%
理学・工学	345,536	204,624	59.2%	3,439	1.0%
複合領域	18,102	5,895	32.6%	94	0.5%

＊予稿集 表3

32

ここまでで機関リポジトリ収録状況はグラフにできるほどの割合ではないことがわかっていただけたかと思うので、分野との関係は表を見ながらお話します。
まずこれは人文・社会、生命科学、理学・工学、複合領域の4領域別に、機関リポジトリへの収録が許可された論文の割合と、実際の収録状況を見たものです。

領域別機関リポジトリ収録状況

領域	総論文数	収録許可論文数	収録許可率	リポジトリ収録数	リポジトリ収録率
人文・社会	147,221	49,168	33.4%	1,415	1.0%
生命科学	484,580	182,299	37.6%	3,570	0.7%
理学・工学	345,536	204,624	59.2%	3,439	1.0%
複合領域	18,102	5,895	32.6%	94	0.5%

＊予稿集 表3

33

論文数の少ない複合領域を除くと、機関リポジトリへの収録が認められた論文の数が最も多いのは理学・工学分野でした。しかし実際の収録率は1%とそれほど高くはありません。
また、人文・社会と生命科学では生命科学の方が若干、収録が許可された論文の割合は高いのですが、実際の収録率は人文・社会領域の方が高くなっていました。

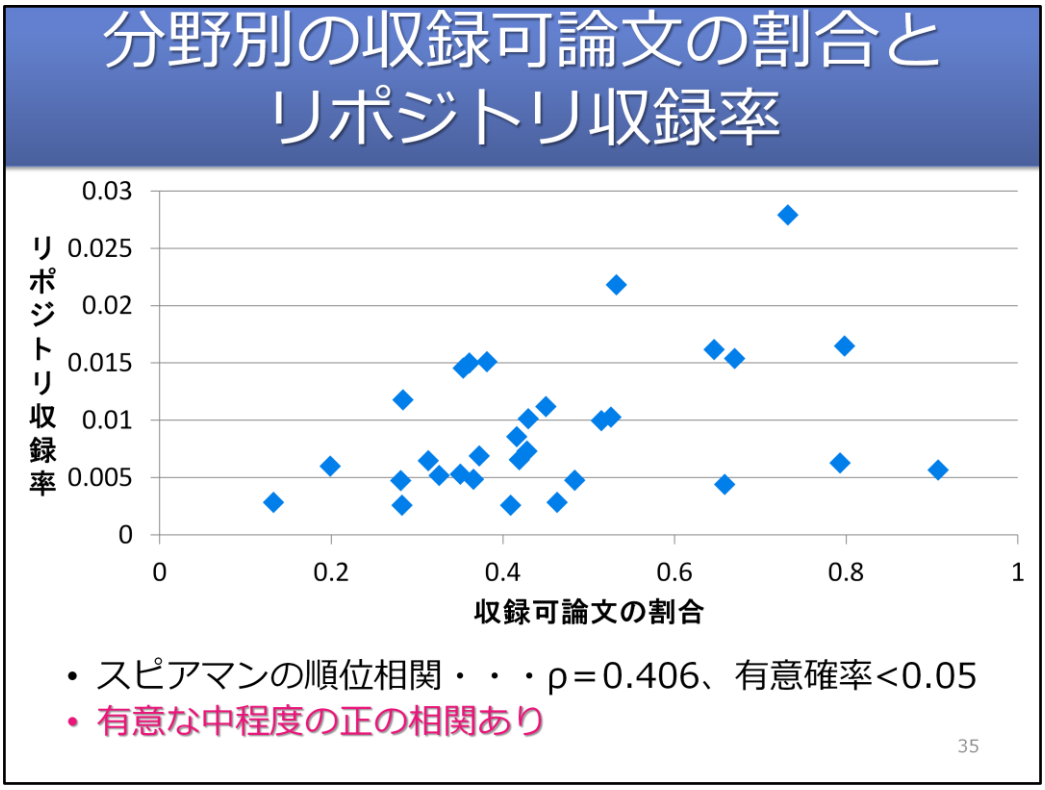
隣接分野間の収録率の違い

- 工学系：
  - 機械工学：収録率2.8%
  - 電気電子工学：1.6%
  - 総合工学：0.4%
  - 土木工学・建築学：0.3%
- 生物系：
  - 基礎生物学：1.6%
  - 総合生物学：0.6%

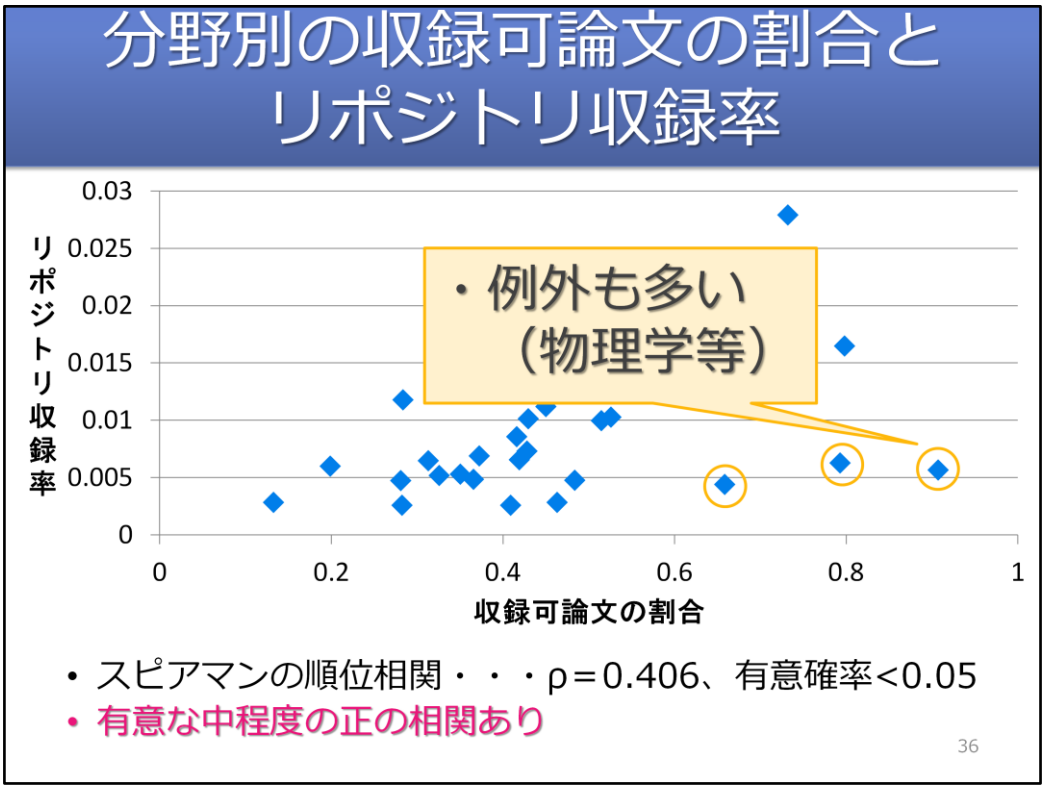
＊詳細は予稿集 表4参照

34

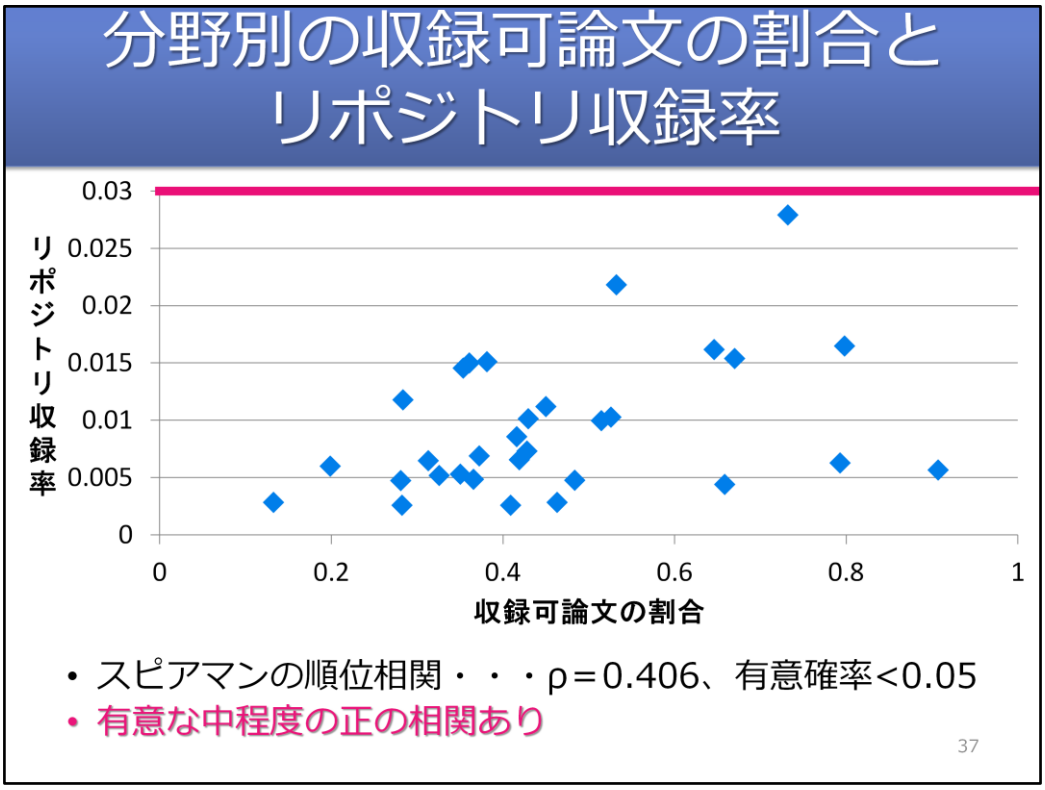
ただし、同じ領域、あるいは隣接分野ですら、機関リポジトリへの論文収録状況は大きく異なります。
例えば最もリポジトリ収録論文の割合が高いのは機械工学で、ほかに電気電子工学でも収録率が高い一方で、同じ工学でも総合工学や土木工学ではごくわずかしか収録されていません。
生物学でも同様で、基礎生物学と総合生物学で傾向は大きく違いました。



また、分野ごとのリポジトリ収録が許可された論文の割合と、実際の収録率の相関をとってみたのがこちらの散布図です。
図からも相関がありそうなことが読み取れますが、相関係数0.406と有意な中程度の正の相関関係にあり、収録できる論文の割合が高ければ実際の収録率も高い傾向がありました。



ただし例外もあります。もっとも収録許可率が高いのは物理学、この右端の点なのですが、その収録率は0.5%程度と全分野の中でも低い方になります。



また、そもそも最も収録率の高い分野ですら3%を割っており、高い・低いを論じる段階にないとも言えます。

分野との関係

- 著作権ポリシーと独立に、機関リポジトリ収録状況には分野による差が存在
- 最も収録数の多い分野でも収録論文は約3%にとどまる
- 突出して収録数の多い分野はない

38

全体に、著作権ポリシーによる差とは別に論文の機関リポジトリ収録状況には分野による差があると言えます。
ただし最も収録率が高い分野でも論文全体の約3%が収録されるにとどまっており、特定の分野で極めて高い、というような傾向はありませんでした。

目次

1.はじめに：研究背景と目的

2.方法・対象

3.分析結果

4.考察と今後の課題

39

最後に分析結果に基づく考察と今後の課題について述べます。

機関リポジトリへの論文収録

- 日本の学協会誌掲載論文中、機関リポジトリに収録されているのはごく僅か
- 2000-2009年発表論文の1.8%
- 学内学会を除けば0.9%

40

本研究の結果から、日本の学協会が発行する雑誌に掲載された論文のうち、機関リポジトリに収録されているものはごく僅かであることがわかりました。
過去10年間に発表された論文の1.8%が機関リポジトリに収録されているに過ぎず、学内学会を除けばその割合は0.9%まで下がります。

著作権ポリシーとの関係(1)

- 著作権ポリシーと機関リポジトリへの論文収録には有意な関係がある
  - 収録可のポリシーを持つ論文：1.3%がリポジトリ収録
  - 収録不可の論文：0.5%
  - 分野別でもポリシーと収録率に相関
- 著作権ポリシーの明示は重要

41

学協会の著作権ポリシーとの関係を見ると、機関リポジトリへの収録を許可しているか否かと論文のリポジトリ収録の間には有意な関係がありました。収録可のポリシーを持っている論文の方がよく収録されており、また収録可能な論文の割合が高い分野ほど、実際にリポジトリに収録されていることもわかりました。
著作権ポリシーの明示は重要であり、SCPJデータベースのような試みは意義のあるものと言えます。

著作権ポリシーとの関係(2)

- 収録可の論文でも実際の収録は僅か
- 学協会が許可を出しても収録されるわけではない
- 「収録可」とされたものを如何に収録していくか？

42

ただしすでに見たように収録が許されている論文でも、実際にリポジトリに収録されているものは1.3%とわずかです。
学協会がリポジトリ収録を許可したからといって収録されているわけではなく、多くの論文が収録されないままです。
これらの論文を如何に機関リポジトリに収録していくかは、引き続き重要な課題と言えるでしょう。

研究上の課題

- 機関リポジトリ以外での提供状況
  - J-STAGEやNII-ELS等
  - CiNii APIで調査可能
- 機関リポジトリでの公開との関係
- 日本の学協会誌掲載論文の電子的利用可能性の全体図

43

最後に研究上の今後の課題として、本研究では機関リポジトリでの収録状況のみ確認したのですが、対象論文の中にはNII-ELS等、その他のサービスによって電子的に提供されているものも多いと考えられます。
その提供状況もCiNIIのAPIで取得することができ、今後はこれらも含めた調査の必要があります。
それにより、例えばNII-ELSで読めるものは機関リポジトリでは公開されていない、といったような相互の関係を明らかにでき、ひいては日本の学協会誌掲載論文が、どのような経路によって電子的に提供されているか、その全体像も描けるのではないかと考えています。

参照文献

[1] “Directory of Open Access Repositories”. http://www.opendoar.org/（2012年4 月30日参照）

[2] “機関リポジトリ一覧”. http://www.nii.ac.jp/irp/list/（2012年4月30日参照）

[3] “IRDBコンテンツ分析システム”. http://irdb.nii.ac.jp/（2012年4月30日参照）

[4] 佐藤翔；大向一輝；関戸麻衣；逸村裕：「アクセスログに基づくCiNiiによる本文提供とその利用状況の分析」，2012年度日本図書館情報学会春季研究集会，三重，2012-05-12.

[5] 佐藤翔；逸村裕：「非学術的活動におけるオープンアクセス文献の活用：機関リポジトリ収録文献のリンク分析」，図書館情報メディア研究，Vol.9，No.1，pp.51-64，2011.

[4] 国立大学図書館協会 学術情報委員会 学術機関リポジトリワーキンググループ：「「学術機関リポジトリに関する調査」報告書」，52p，2010，http://wwwsoc.nii.ac.jp/anul/j/publications/reports/repository1.pdf（2012年4月30日参照）

[5] “学協会著作権ポリシーデータベース”. http://scpj.tulips.tsukuba.ac.jp/（2012年4月30日参照）

[6] Harnad, Stevan: “Estimating Japan's Annual Rate of Journal Article Self-Archiving”, Open Access Archivangelism, 2010, http://openaccess.eprints.org/index.php?/archives/763-Estimating-Japans-Annual-Rate-of-Journal-Article-Self-Archiving.html（2012年4月30日参照）

発表は以上です、ありがとうございました。
主要参考文献はこちらに示したとおりです。