

日本におけるオープンアクセスとオープン・サイテーションの現状

西岡 千文（京都大学附属図書館）

2019年5月20日（月） 2019年度京都大学図書館機構講演会



京都大学
KYOTO UNIVERSITY



発表内容

- 背景：引用データのオープン化とその重要性
- データセット：分析に使用するデータ
- 結果：オープン・サイテーションとオープンアクセスの現状
- まとめ

発表内容

- 背景: 引用データのオープン化とその重要性
- データセット: 分析に使用するデータ
- 結果: オープン・サイテーションとオープンアクセスの現状
- まとめ

引用データのオープン化

引用データは、研究評価、研究プロセスの理解、図書館の蔵書形成など様々な目的で利用されている [Smith 81]

しかし、機械可読なアクセスの実現は遅れている

- 複雑なライセンスでの保護
- 引用データの整備のためのリソース不足（予算、人材）

近年、引用データのオープン化が進展している

- **I4OC (Initiative for Open Citations)**
- Microsoft Academic Graph
- Springer Nature SciGraph

[Smith 81] Smith, L. C.: Citation analysis, Library Trends, Vol. 30, No. 1, pp. 83–106 (1981).

オープンとは？

様々な「オープン」

★	Available on the web (whatever format) <i>but with an open licence, to be Open Data</i>
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

<https://www.w3.org/DesignIssues/LinkedData.html>



学术论文のオープンアクセスは多くの場合★に該当する。対して、引用データのオープン化は★ ★ ★ ★や★ ★ ★ ★ ★を目指す。

引用データの重要性

引用データのオープン化の取り組み開始から数年しか経っていないので、
学術情報流通に与えた影響を評価することは難しいが…

- ◆ 電子図書館では、「文献のページにおける引用文献・被引用文献の閲覧」という情報検索行動は、クエリ検索など他の情報検索行動と比較すると、表示された文献に肯定的な評価を示す行動をより多く引き起こす [Kacem et al. 18]

引用データのオープン化が学術情報流通や研究評価に及ぼす影響は大きくなることが想定されるため、
日本の学術出版物におけるオープン・サイテーションの現状を調査

[Kacem et al. 18] Kacem, A. and Mayr, P.: Analysis of search stratagem utilisation, Scientometrics, Vol. 116, No. 2, pp. 1383– 1400 (2018).

発表内容

- 背景：引用データのオープン化とその重要性
- データセット：分析に使用するデータ
- 結果：オープン・サイテーションとオープンアクセスの現状
- まとめ

引用データ

- I4OCはCrossRefのメタデータに引用文献を登録してオープンにすることを推奨
- I4OCの取り組みによってオープンになったデータをOpenCitationsが機械可読なフォーマットで公開

OpenCitations CorpusからCOCI (Open Citations Index of CrossRef Open DOI-to-DOI references) (2018年11月22日版)の利用

- データセット詳細
 - 引用元となる文献についてはCrossRef DOIをもつ文献のみが対象
 - 引用先としてはあらゆるDOIをもつ文献が対象
 - 46,530,436 件の文献間の引用関係が 449,842,375 件収録

日本の学術出版物 [1/3]

- 本発表では、日本の学術出版物は日本の出版者から公開された出版物とする
 - 日本の学術出版物≠日本の著者による出版物
 - オープンアクセスとは異なり、引用データをオープンにする主体は著者ではなく出版者であることが多い

JaLCメタデータ(2018年9月7日版)の利用

- JaLCによって付与されたDOIをもつ6,370,356件の文献
- COCIで引用元として収録されている文献(引用文献リストを公開している文献)はCrossRef DOIをもつ文献のみで、JaLC DOIが付与されたものは、引用元として収録されない



日本の学術出版物 [2/3]

引用文献を公開している文献を分析する際には、JaLC経由で登録された CrossRef DOIを保有する論文を対象とする必要

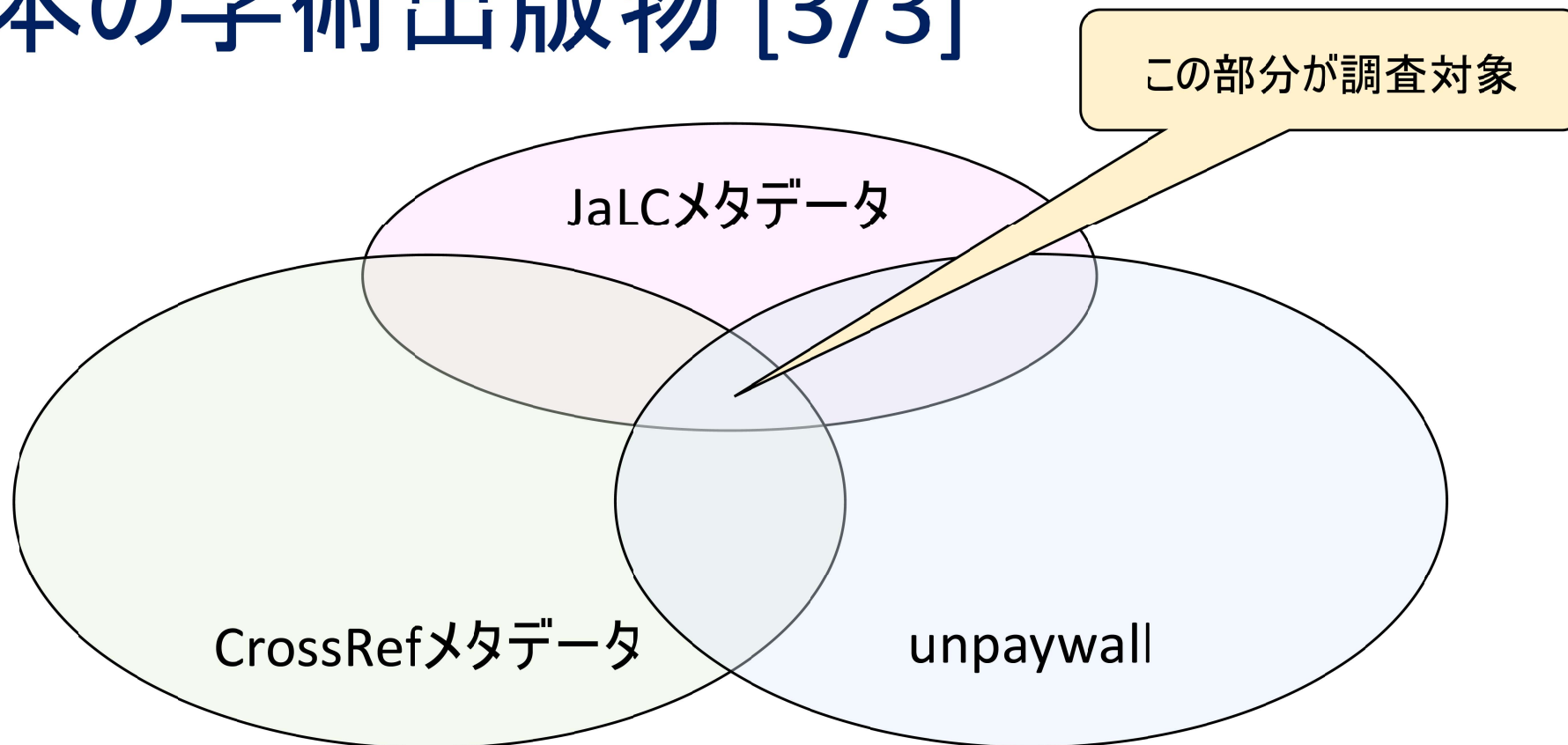
「JaLC経由で登録されたCrossRef DOIをもつ文献」
=「JaLCメタデータの収録文献」∩「CrossRefメタデータの収録文献」

さらに...「各文献のオープンアクセス状況と絡めて解析したい」

「調査対象文献」=「JaLC経由で登録されたCrossRef DOIをもつ文献」
∩「unpaywallの収録文献」

- unpaywallはあらゆる文献の合法的にオープンアクセスとなっている版を探索・提供するウェブブラウザの拡張機能
- 拡張機能が利用するデータベースは公開
 - CrossRef DOIをもつ全文献(99,940,229件)のDOIならびにオープンアクセスに関する情報等を収録

日本の学術出版物 [3/3]



JaLCメタデータに収録されている6,370,356件のうち、
2,049,891件 (32.18%)が調査対象

CrossRefメタデータ提供元: https://archive.org/download/crossref_doi_dump_201809

unpaywall: <https://unpaywall.org/products/snapshot>

JaLCメタデータ: https://japanlinkcenter.org/top/material/material_metadata.html

引用データの公開状況の判定

- **Open (公開)** : 文献がCOCIに引用元として収録(引用文献がCrossRefに登録されており公開)
 - **Closed (非公開)** : CrossRefの“**reference-count**”フィールドが1以上でCOCIに引用元としての収録なし(引用文献がCrossRefに登録されており、非公開)
 - **その他** : 下記の2ケースが含まれるが、判別方法がない
 - **No existence (引用文献なし)** : そもそも引用文献が存在しない(編集後記、目次等)
 - **Incomplete (不完全)** : 引用文献は存在するが、CrossRefに登録されていない
-
- **引用先として公開** : 文献がCOCIに引用先として収録
 - **オープンアクセス** : unpaywallでの“**is_oa**”フィールドがtrue

文献の分類

- 分類ごとの引用データのオープン化状況についても調査
- NDLの雑誌記事索引採録誌一覧と文献のISSNから収録雑誌を特定
- 分類として、雑誌記事索引採録誌一覧で各雑誌に付与されている国立国会図書館分類表の分類を利用
- JaLC メタデータの6,370,356件の文献のうち, 4,788,310 件(75.17%)の文献がISSN を保有
- 3,099,143 件(48.65%)が雑誌記事索引採録誌一覧と紐付けられ, 3,089,698 件(48.50%) が分類を保有

発表内容

- 背景：引用データのオープン化とその重要性
- データセット：分析に使用するデータ
- 結果：オープン・サイテーションとオープンアクセスの現状
- まとめ

世界の状況との比較

	全件数	引用文献を公開 (%)	引用文献を非公開 (%)	引用先として公開 (%)	オープンアクセス (%)
日本	2,049,891	386,632 (18.86)	148,130 (7.23)	539,584 (26.32)	1,696,636 (82.77)
世界	99,848,571	24,178,446 (24.22)	16,589,545 (16.61)	38,481,195 (38.54)	24,961,752 (25.00)

- 日本の引用データのオープン化の割合は低い
 - しかし非公開にしている割合は低い
 - 引用文献を組織化して、CrossRefに登録することが課題である
- 日本の学術出版物は引用先としての収録が少ない
 - 日本の学術出版物の多くは日本語であるため、世界の学術出版物よりも読者は少ない
- 日本の学術出版物は学協会によるものが大きな割合を占めるため、オープンアクセスの割合は圧倒的に高い

オープンアクセス

オープンアクセス(OA)である文献とそうでない文献のあいだに、引用データのオープン化傾向に差異があるか？

	全件数	引用文献を公開	引用文献を非公開	引用先として公開
オープンアクセスである文献	1,696,636	316,810 (18.67%)	124,015 (7.31%)	475,950 (28.05%)
オープンアクセスでない文献	353,255	69,822 (19.77%)	24,115 (6.83%)	63,634 (18.01%)
計	2,049,891	386,632 (18.86%)	148,130 (7.23%)	539,584 (26.32%)

- 引用文献の公開においては、OAでない文献が若干多いが、あまり差はない
- 引用先としてのオープンでは、OAである文献が多い
 - 発見可能性の高さや引用回数多さといったオープンアクセス文献の特長についての結果と合致

DOAJへの収録

DOAJ (Directory of Open Access Journals): オープンアクセス学術誌とその論文をリストしたウェブサイトである。リストされる学術誌は精査される。

DOAJに収録されている文献とそうでない文献のあいだに、引用データのオープン化傾向に差異があるか？

	全件数	引用文献を公開	引用文献を非公開	引用先として公開
DOAJ収録文献	3,760	1,873 (49.81%)	992 (26.38%)	2,858 (76.01%)
DOAJ非収録文献	2,046,131	384,759 (18.80%)	147,138 (7.19%)	536,726 (26.23%)
計	2,049,891	386,632 (18.86%)	148,130 (7.23%)	539,584 (26.32%)

- **そもそもDOAJ収録の日本の学術誌の文献が少ない**
- DOAJ収録文献は約半数の文献で引用文献を公開している一方、約4分の1は非公開にしている
 - 約半数という数値は世界の文献を対象とした分析での数値(48.99%)とほぼ一致

分類 [1/3]

文献の分類のあいだに、引用データのオープン化傾向に差異があるか？

	全件数	引用文献を公開 (%)	引用文献を非公開 (%)	引用先として公開 (%)	オープンアクセス (%)
ZA 政治・法律・行政	4,735	100 (2.10)	55 (1.16)	6 (0.12)	4,487 (94.76)
ZD 経済	10,179	1,534 (15.07)	1,675 (16.46)	1,012 (9.94)	9,213 (90.50)
ZE 社会・労働	10,759	1,446 (13.44)	1,367 (12.71)	2,363 (21.96)	10,193 (94.74)
ZF 教育	17,621	3,234 (18.35)	1,992 (11.30)	2,541 (14.42)	16,749 (95.05)
ZG 歴史・地理	14,740	192 (1.30)	990 (6.72)	1,278 (8.67)	13,670 (92.74)
ZH 哲学・宗教	15,104	1 (0.01)	124 (0.82)	41 (0.27)	14,289 (94.60)
ZK 芸術	6,162	595 (9.66)	605 (9.82)	1,102 (17.88)	5,821 (94.47)
ZM 科学技術一般	164,958	28,172 (17.08)	11,250 (6.82)	45,866 (27.80)	151,620 (91.91)
ZN 建設工学・機械工学	248,957	34,153 (13.72)	26,391 (10.60)	51,123 (20.53)	221,177 (88.84)
ZP 化学・化学工業	364,818	95,302 (26.12)	27,415 (7.51)	138,717 (38.02)	322,885 (88.51)
ZR 生物学	192,386	35,434 (18.42)	17,268 (8.98)	79,911 (41.54)	179,536 (93.32)
ZS 人類学・医学・薬学	407,762	68,204 (16.73)	21,191 (5.20)	108,246 (26.55)	374,898 (91.94)
不明	591,711	118,267 (19.99)	37,807 (6.39)	107,381 (18.15)	372,099 (62.89)
計	2,049,891	386,632 (18.86)	148,130 (7.23)	539,584 (26.32)	1,696,636

分類 [2/3]

- 引用元・引用先としてのオープンに注目した場合、STM分野、特に化学・化学工業に属する文献で、引用データのオープン化が進展している
 - 同一分野の文献を引用する機会が多いので、引用元としてオープンになると、その分野の引用先としてのオープンも進展する
 - 引用文献を非公開にしている文献も7～9%存在する

分類 [3/3]

- 対して、歴史・地理、哲学・宗教など人文学系分野においては、引用データのオープン化は遅れている
 - 引用文献を公開している文献より非公開の文献の方が多い
 - 人文学系の文献は他の分野と比較すると引用されにくい[Althouse et al. 09]という調査結果を反映している
 - 同一分野の文献を引用する機会が多いので、引用元としてオープンにならないと、引用先としてもオープンにならない

オープンデータで研究評価がされるようになると、分野が過小評価されかねない

- いずれの分野でもオープンアクセス率は約9割で、分野を特定できなかったもので約6割

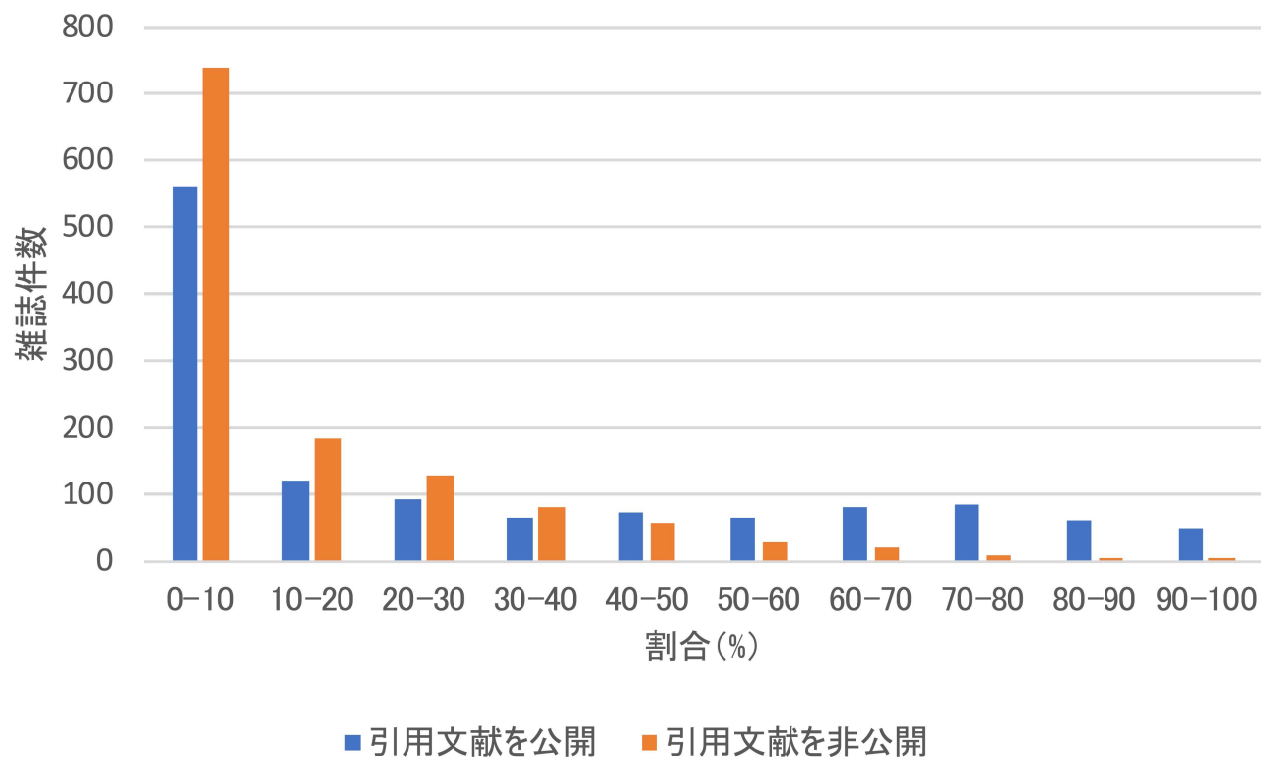
[Althouse et al. 09] Althouse, B. M., West, J. D., Bergstrom, C. T. and Bergstrom, T.: Differences in impact factor across fields and over time, Journal of the American Society for Information Science and Technology, Vol. 60, No. 1, pp. 27–34 (2009).

収録雑誌 [1/3]

- 文献のメタデータの付与・登録方法は、各雑誌の収録文献のあいだでは概ね統一されている
- 引用文献のCrossRefメタデータへの登録やそれらの公開については、出版者や編集者の意向が反映される

収録雑誌のあいだに、引用データのオープン化傾向に差異があるか？

収録雑誌 [2/3]



ほとんどの雑誌で引用文献をオープンにしている割合は10%以下である

収録雑誌 [3/3]

最も引用文献を公開している収録雑誌

雑誌名	割合 (%)
BIOPHYSICS	100.00
Polymer Journal	99.58
MICROBIOLOGY and IMMUNOLOGY	99.55
Chemistry Letters	99.08
MATERIALS TRANSACTIONS	98.41
Circulation Journal	98.33
Materials Transactions, JIM	97.88
Journal of the Physical Society of Japan	97.53
Journal of Oral Biosciences	97.41
Journal of Pharmacological Sciences	97.38
Bulletin of the Chemical Society of Japan	96.90

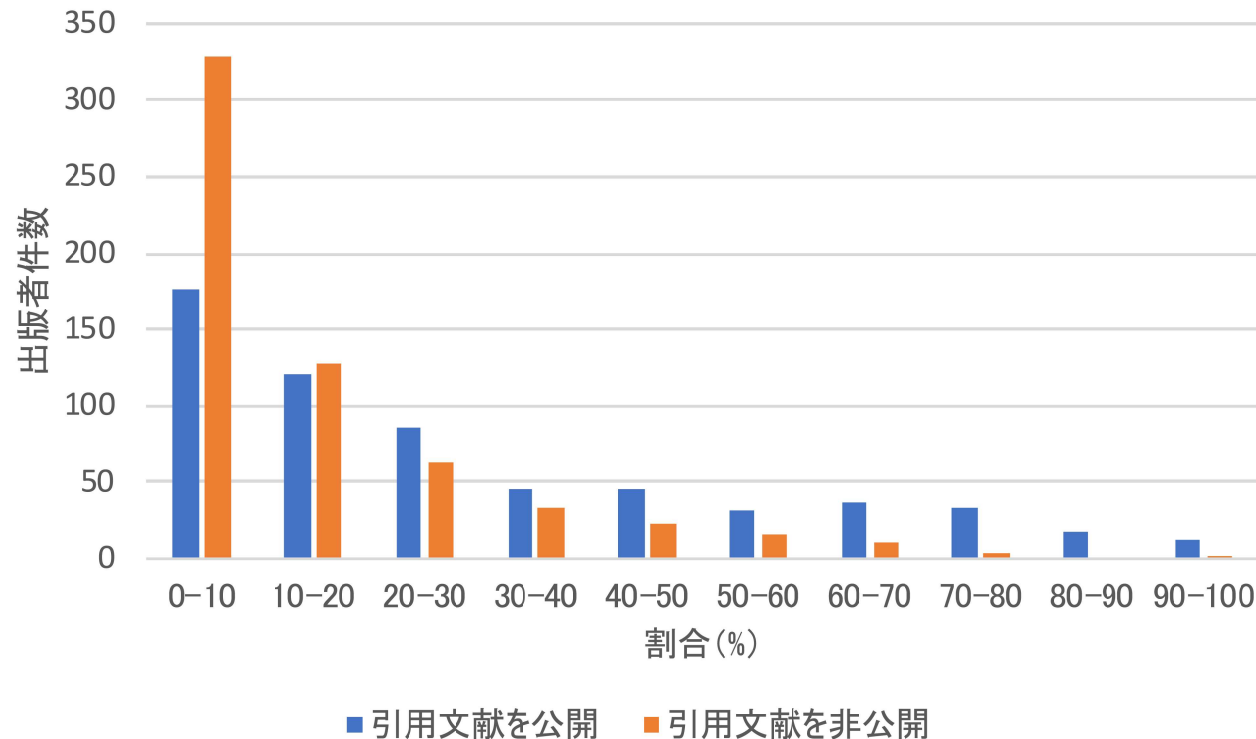
収録文献が100件以上ある雑誌で、引用文献をオープンにしている割合が最も高い雑誌10誌

- STM分野に属する英文誌
- WoS Core Collectionに収録されていてIFを保有していることから国際的な認知高
- 日本の学術出版物の多くがJ-STAGEや機関リポジトリを出版プラットフォームとして利用していることに対して、10誌中7誌がNature、Science Directといった商業出版社のプラットフォームを利用

商業出版社の慣習が引用文献の登録・公開に影響を与えている？

出版者 [1/3]

異なる出版者の学術出版物のあいだに、引用データのオープン化傾向に差異があるか？



雑誌同様、ほとんどの出版者で引用文献をオープンにしている割合は10%以下である

出版者 [2/3]

大学からの出版物

	全件数	引用文献を公開 (%)	引用文献を非公開 (%)	引用先として公開 (%)	オープンアクセス (%)
Tohoku University Medical Press	10,426	2151 (20.63)	426 (4.09)	4504 (43.20)	10066 (95.55)
Department of Anatomy Keio University School of Medicine	2,651	276 (10.41)	407 (15.35)	1282 (48.36)	2513 (94.79)
Mathematical Institute, Tohoku University	2,551	0 (0.00)	0 (0.00)	1245 (48.80)	2423 (94.98)
Nihon University School of Dentistry	1,994	727 (36.46)	242 (12.14)	1214 (60.88)	1889 (94.73)
The University of Occupational and Environmental Health, Japan	1,679	126 (7.50)	73 (4.35)	346 (20.61)	1602 (95.41)
The Medical Society of Kansai Medical University	1,516	23 (1.52)	4 (0.26)	35 (2.31)	1446 (95.38)
Faculty of Mathematics, Kyushu University	1,358	527 (38.81)	780 (57.44)	387 (28.50)	1293 (95.21)
The Showa University Society	906	642 (70.86)	53 (5.85)	97 (10.71)	847 (93.49)
University of Tokushima Faculty of Medicine	742	567 (76.42)	141 (19.00)	527 (71.02)	698 (94.07)
The Nihon University Medical Association	677	509 (75.18)	115 (16.99)	35 (5.17)	655 (96.75)
Nihon University School of Dentistry at Matsudo Research Institute of Oral Science	424	366 (86.32)	54 (12.74)	124 (29.25)	394 (92.92)
Graduate School of Information Sciences, Tohoku University	415	194 (46.75)	69 (16.63)	131 (31.57)	397 (95.66)

出版者 [3/3]

	全件数	引用文献を公開 (%)	引用文献を非公開 (%)	引用先として公開 (%)	オープンアクセス (%)
Yamaguchi University Medical Association	300	157 (52.33)	129 (43.00)	38 (12.67)	287 (95.67)
Showa University Dental Society	212	75 (35.38)	50 (23.58)	15 (7.08)	198 (93.40)
Kyushu University	95	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Institute of Human Culture Studies, Otsuma Womens University	89	22 (24.72)	44 (49.44)	6 (6.74)	88 (98.88)
Marianna University Society of Medical Science	69	59 (85.51)	6 (8.70)	8 (11.59)	66 (95.65)
Institute for Policy Studies and Social Innovation, University of Hyogo	46	21 (45.65)	16 (34.78)	7 (15.22)	45 (97.83)
Graduate School of Environmental Science, Okayama University	36	26 (72.22)	10 (27.78)	11 (30.56)	34 (94.44)
Kyoto University	30	0 (0.00)	0 (0.00)	0 (0.00)	30 (100.00)
Research Institute for Economics and Business Administration - Kobe University	30	24 (80.00)	6 (20.00)	15 (50.00)	29 (96.67)
Research Institute for Mathematical Sciences, Kyoto University	2	0 (0.00)	0 (0.00)	0 (0.00)	2 (100.00)

ほとんどの大学でほぼ全ての出版物がオープンアクセス
→けれども引用文献が非公開になっているものも多く存在

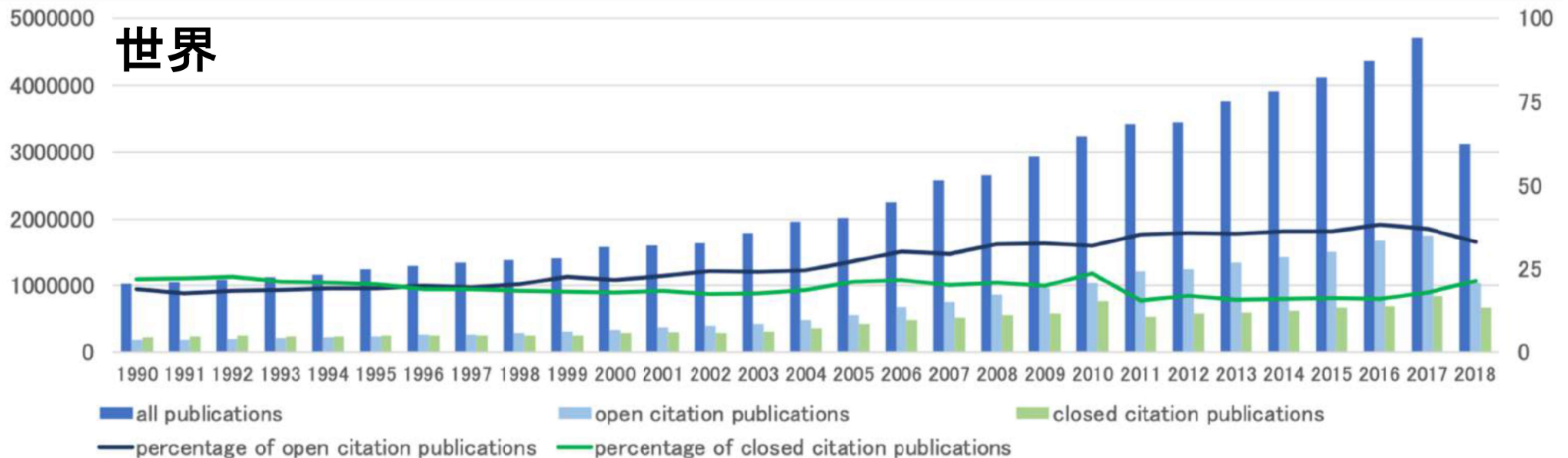
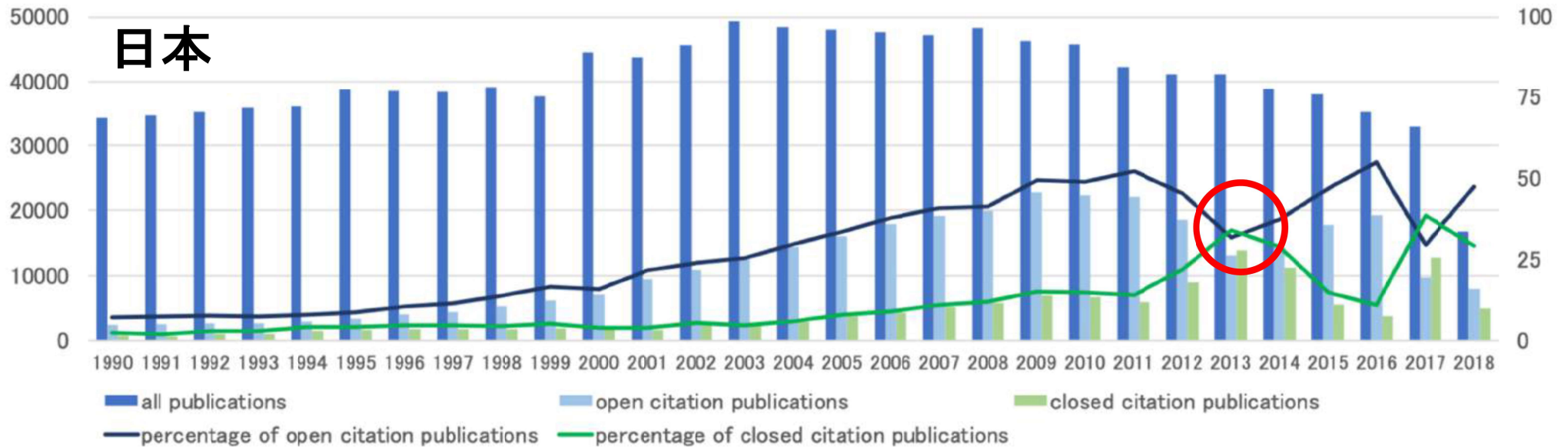
出版年 [1/3]

- 引用データは著作権などで保護されているという考えもあるため、論文と同様に、出版社が定めるエンバゴ期間にはオープンにできないということも考えられる
- また、古い論文では、引用文献がメタデータに整備されていないということも考えられる

異なる出版年の文献のあいだに、引用データのオープン傾向に差異があるか？

出版年 [2/3]

直近の10年間に出版された文献に注目すると、概ね40%以上の文献の引用文献がオープンになっているが過去の文献では、オープン化は進んでいない



出版年がオープン化の割合に与える影響は少ない

出版年 [3/3]

- 近年では、日本の学術出版物における引用文献のオープン化は海外と比較して進んでいるといえる
 - しかし、近年にいくほど非公開の割合も増えている
- 過去の文献については、公開の割合も非公開の割合も低いので、引用データの整理が課題である
 - 過去の文献が閲覧される回数は比較的少ないが、研究の検証可能性の向上のためにも、過去の文献の引用文献を後悔することは重要
- 世界の学術出版物における近年の引用文献のオープン化の割合が低い理由として、大手商業出版社による影響が挙げられる
 - Elsevier が出版元である文献の引用文献をオープンにするだけで、割合は10%増加

紀要 [1/2]

- 出版者のプラットフォームでは各文献のページにて引用文献や被引用文献へのリンクが付与されている
 - 文献に述べられている科学的知見の検証可能性の向上
- 多くの機関リポジトリでは引用文献・被引用文献へのリンクは付与されていない
- 紀要は機関リポジトリがプライマリな公開場所として利用されていることが多い [竹内 12]
- 機関リポジトリで引用文献・被引用文献へのリンクを表示するにあたっては、引用データがオープンになっていることが望ましい

紀要文献に注目して、引用データのオープン化の現状を分析する

[竹内 12] 竹内比呂也: 大学紀要というメディア: 限りなく透明に近いグレイ? (<特集>灰色文献), 情報の科学と技術, Vol. 62, No. 2, pp. 72-77 (2012).

紀要 [2/2]

- 紀要に収録されている文献：18,079 件
 - 国立国会図書館分類ZV1(紀要)をもつ
 - 雑誌記事索引採録誌一覧では、23,910誌のうち3,029誌にZV1の分類が与えられている
 - 47誌の紀要のうちいずれかに収録されている
- 紀要に収録かつCrossRef DOIをもつ文献：6,533 件
 - 日本學士院紀要、Proceedings of the Japan Academy、Synthesiology、社会技術研究論文集、人間環境学研究の5誌のうちいずれかに収録されている

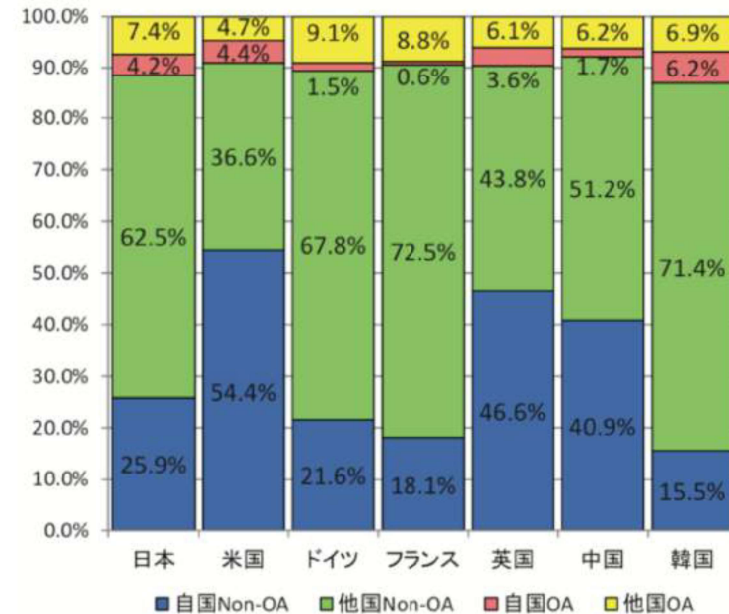
2,131件 (32.62%) が引用元としてオープン
(サンプル数が少ないので、Concreteな結論ではない)

本調査の対象となった文献

種別	割合(%)
国内のOA誌	4.2
国内のNon-OA誌	7.4
国外のOA誌	25.9
国外のNon-OA誌	62.5

このうちの約30%が今回の調査対象

(B) 割合



(注)年は論文の出版年を使用している。雑誌の種類は Journal、論文の種類は Article、Conference Paper、Review である。

福澤尚美. ジャーナルに注目した主要国の論文発表の特徴: オープンアクセス、出版国、使用言語の分析. 研究・イノベーション学会年次学術大会講演要旨集, vol.31, pp.624-627.

実際のOA率



Daniel Hook, Mark Hahnel, Ian Calvert. The Ascent of Open Access. 2019.

ジャパンリンクセンター (JaLC)、「JaLCメタデータ」の提供を開始

Posted 2017年3月1日

2017年3月1日、ジャパンリンクセンター (JaLC) が、「JaLCメタデータ」の提供を開始しています。

「JaLCメタデータ」には、JaLC ウェブサイトで公開している、DOIと書誌データ（論文タイトル、著者名、出版社名、ISSN、巻、号、開始ページ）が含まれ、LOD (Linked Open Data) 形式で構成されており、圧縮ファイル（tar.gz形式）で提供されています。

データは、複製、公衆送信、翻訳・変形等の翻案等、自由に利用でき、商用利用も可能とのことです。

JaLC 新着情報

<https://japanlinkcenter.org/top/index.html>

※「2017年03月01日 各種資料 「JaLCメタデータ」の提供を開始しました。」とあります。

JaLCメタデータ (JaLC)

https://japanlinkcenter.org/top/material/index.html#material_metadata

ジャパンリンクセンター公開書誌データ等利用規約 (JaLC, 2017/3/1付け)

https://japanlinkcenter.org/top/doc/JaLC_shosidata_riyoukiyaku.pdf

カレントアウェアネス-R, 学術情報流通, 日本

参照(1824) [印刷用ページ](#)

<http://current.ndl.go.jp/node/33576>

JaLCからの引用文献公開への期待

3. 公開するメタデータの範囲

オープン化を検討しているメタデータは、以下のとおりである。

- ・ 書誌データ（タイトル、著者、ジャーナル、巻、号、開始ページ等）
- ・ 引用文献
- ・ 被引用情報
- ・ 抄録

抄録以外のメタデータは、事実を記載したものであることから著作権が発生しないため、一律オープン化することとする。

一方、抄録については著作権が発生するため、オープン化するためには著作権者に許諾を得る必要があると考えている。

ジャパンリンクセンター運営委員会. メタデータのオープン化について. 2019.

発表内容

- 背景：引用データのオープン化とその重要性
- データセット：分析に使用するデータ
- 結果：オープン・サイテーションとオープンアクセスの現状
- まとめ

まとめ

- 世界の学術出版物の引用文献のオープン化の割合が24.22%であるのに対して日本の学術出版物は18.86%であり、海外と比較すると進んでいない
- 過去10年に注目すると、日本の引用文献のオープン化の割合は海外と比較して高い
 - しかし、非公開の割合も高くなっている
- オープンアクセスである文献でも引用文献を非公開にしている文献も存在する
- 日本の学術出版物においては、STM分野、特にそれらの分野の英文誌が引用文献のオープン化の割合が高い
- 人文学系分野での引用データのオープン化の割合は総じて低いので、これらの分野における引用文献のオープン化への認知の向上と引用データの組織化の支援が必要

終

参考文献:西岡 千文, 亀田 堯宙, 佐藤 翔. 日本の学術出版物における引用データのオープン化の現状分析. 研究報告人文科学とコンピュータ(CH), vol.2019-CH-120, no.5, pp.1-8, 2019.