

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

西岡 千文^{1,a)} 亀田 堯宙^{2,b)} 佐藤 翔^{3,c)}

概要: 近年、学術出版物のオープンアクセスが進展し、自由にアクセス可能な学術情報が蓄積されている。一方で、研究評価など様々な目的で利用されている学術出版物の引用データに関しては、機械可読なアクセスのオープン化が遅れてきた。このような状況を解決するために、I4OC (Initiative for Open Citations) が学術機関と出版社によって設立された。本稿では、I4OC が公開している引用データを、JaLC メタデータと unpaywall により公開されているデータとともに利用することで、日本の学術出版物の引用データのオープン化の現状分析を試みる。

Current State of Open Citation in Japanese Scholarly Literature

1. はじめに

近年、学術出版物のオープンアクセスが進展し、自由にアクセス可能な学術情報が蓄積されている。Piwowar らの調査 [1] によると、約 28% の学術出版物がオープンアクセスで提供されており、2015 年に出版された学術出版物については約 45% がオープンアクセスで公開されている。一方で、研究評価、研究プロセスの理解、図書館の蔵書形成など [2] 様々な目的で利用されている学術出版物の引用データに関しては、機械可読なアクセスの実現が遅れている。このような状況を解決するために、2017 年 4 月に I4OC (Initiative for Open Citations)^{*1} が学術機関と出版社によって設立された。I4OC は、機械可読な形式で引用データをオープンにすることで、引用データの利用可能性を高める。I4OC 設立前は、CrossRef に登録された引用データの約 1% がオープンであったが、I4OC 設立 1 年後には約 40% に数値が上昇している^{*2}。CrossRef DOI が

付与された全出版物に注目すると、2018 年 9 月時点では 24.20% の文献の引用データがオープンになっている。

なお、本稿における「データのオープン」とは、5 つ星オープンデータモデル^{*3}における 4 つ星あるいは 5 つ星を意味する。5 つ星オープンデータモデルとは、Tim Berners-Lee によって提唱されている、オープンデータを 5 段階で評価するスキームである。4 つ星は RDF など W3C が定める標準に沿ったフォーマットで公開されていることを、5 つ星は他のデータとリンクしていることを指す。文献がオープンアクセスで公開されていれば引用データにも無料でアクセス可能であるが、RDF などの機械可読なフォーマットで公開されていない場合は、オープンであると本稿では判断しない。I4OC は 5 つ星オープンデータを指向して、RDF での引用データの公開や SPARQL エンドポイントを提供している。

執筆時点では I4OC による引用データオープン化の取り組み開始から 2 年しか経過していないため、引用データのオープン化が文献の発見可能性の向上 [3], [4] や引用回数の増加 [1], [3], [4] に貢献したといったオープンアクセスに与えられる肯定的な評価を断定することはできない。しかし、これまでに I4OC によって公開された引用データは学術情報探索プラットフォーム Dimensions^{*4}, Wikidata,

4 日

^{*3} <https://www.w3.org/DesignIssues/LinkedData.html>, 最終アクセス日: 2019 年 4 月 4 日

^{*4} <https://app.dimensions.ai/discover/publication>, 最終

¹ 京都大学附属図書館

Kyoto University Library

² 京都大学東南アジア地域研究研究所

Center for Southeast Asian Studies, Kyoto University

³ 同志社大学免許資格課程センター

Center for License and Qualification, Doshisha University

a) nishioka.chifumi.2c@kyoto-u.ac.jp

b) kameda@cseas.kyoto-u.ac.jp

c) min2fly@slis.doshisha.ac.jp

^{*1} <https://i4oc.org/>, 最終アクセス日: 2019 年 3 月 28 日

^{*2} <https://i4oc.org/press.pdf>, 最終アクセス日: 2019 年 4 月

OpenAIRE^{*5}, LOC-DB[5]などのプロジェクトや計量書誌学など、様々な場面での利活用が進んでいる[6]。また、電子図書館でのユーザの情報検索行動に関する研究[7], [8]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探索行動は、クエリ検索など他の情報探索行動と比較して、PDF ファイルの閲覧、書誌情報のエクスポートといった探索によって得ら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ことが報告されている。よって、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予想される。

本稿では、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を分析する。「日本の学術出版物」とは、日本の著者により執筆された出版物ではなく、日本の出版者から公開された出版物を意味する。これは、学術出版物の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ためである。分析にあたっては、I4OC により公開された引用データを、JaLC メタデータと unpaywall のデータセットとともに利用する。

本稿の構成を以下に示す。2 章では、分析に利用したデータセットについて詳述する。3 章では、出版物の分類、収録雑誌、出版年など様々な観点から、引用データのオープン化の状況について分析結果を示す。4 章を本稿のむすびとする。

2. データセット

本章は、分析に利用したデータを示す。2.1 節では引用データについて、2.2 節では分析対象となる日本の学術出版物についてそれぞれ述べる。本稿では分類や雑誌ごとにも分析を実施するため、2.3 節では出版物の分類・収録雑誌の取得方法を示す。

2.1 引用データ

引用データとして、I4OC が公開するデータセットを利用する。I4OC は二種類の方法で引用データを提供している。一点目は、CrossRef REST API での提供である。CrossRef メタデータには **reference** というフィールドがあり、出版者は出版物のメタデータを登録する際、当該フィールドに出版物が引用している文献を記述できる。当該フィールドが利用されているかつ公開されている場合は、CrossRef REST API^{*6}を利用して各出版物の引用文献を取得できる。二点目は、OpenCitations Corpus^{*7}で

の提供である。OpenCitations では、引用データを RDF や csv など機械可読なフォーマットでデータセットとして公開している。I4OC の活動による CrossRef DOI をもつ出版物の引用データオープン化の進展を受けて、2018 年以降は、COCI OpenCitations Index of CrossRef Open DOI-to-DOI references^{*8}として引用データを公開している。データセットの名称のとおり、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となっている。引用先となる文献については CrossRef DOI を含むあらゆる DOI をもつ文献が対象である。本稿は、執筆時点で最新版（2019 年 11 月 22 日版）の COCI データセット^{*9}を利用する。表 1 にデータセット概要を示す。表 1 にあるとおり、データセットには 46,530,436 件の文献間の引用関係が 449,842,375 件記述されている。なお、これらの数値はウェブサイト上に記述されている数値^{*10}と異なる。本稿では著者らが精査した結果である表 1 の数値を利用する。

なお、オープンになっている引用データとして Microsoft Academic Graph[9] も挙げられるが、引用データの生成過程が比較的不明瞭であることから、本稿では COCI のデータセットを利用する。

表 1 COCI データセット概要

	件数
引用元として収録されている文献	24,182,977
引用先として収録されている文献	38,481,195
収録されている文献	46,530,436
引用関係	449,842,375

2.2 日本の学術出版物

1 章で述べたとおり、「日本の学術出版物」とは日本の出版者から公開された出版物を意図する。本稿では日本の学術出版物を「ジャパンリンクセンター（JaLC）によって付与された DOI をもつ出版物」とであると仮定し、2018 年 9 月 7 日版の JaLC メタデータ^{*11}を利用する。JaLC メタデータには、JaLC によって付与された DOI をもつ文献が 6,370,356 件収録されている。

しかし前節で述べたとおり、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 をもつ文献のみである^{*12}。よって、JaLC

アクセス日：2019 年 4 月 4 日

^{*5} <https://www.openaire.eu/>, 最終アクセス日：2019 年 4 月 4 日

^{*6} <https://www.crossref.org/services/metadata-delivery/rest-api/>, 最終アクセス日：2019 年 4 月 4 日

^{*7} <http://opencitations.net/>, 最終アクセス日：2019 年 3 月 28 日

^{*8} 以下、特段の断りがなければ COCI は COCI OpenCitations Index of CrossRef Open DOI-to-DOI references を意味する。

^{*9} <https://doi.org/10.6084/m9.figshare.6741422.v3>, 最終アクセス日：2019 年 3 月 28 日

^{*10} <http://opencitations.net/download>, 最終アクセス日：2019 年 4 月 5 日

^{*11} https://japanlinkcenter.org/top/material/material_metadata.html, 最終アクセス日：2019 年 4 月 5 日

^{*12} 一件の例外が観察された。文献 (<https://doi.org/10.1589/jpts.30>) は、JaLC DOI が付与されているが、COCI にて引用元の文献として収録されていた。

表 2 日本の学術出版物データセット概要

出版物	全件数	分類をもつ文献件数
JaLC メタデータに収録されている出版物	6,370,356	3,036,483
unpaywall に収録されている出版物 (≈JaLC 経由で登録された CrossRef DOI をもつ文献)	2,049,903	1,458,192
引用元として COCI に収録されている出版物	386,635	268,367
unpaywall で OA と判断されている出版物	1,696,647	1,324,548
引用先として COCI に収録されている出版物	544,259	435,433
COCI に収録されている出版物	714,903	549,809

DOI が付与された学術出版物については、引用文献のデータは COCI に収録されていない。

引用文献を公開している文献に注目して分析する際には、JaLC 経由で登録された CrossRef DOI を保有する論文を対象とする^{*13}。JaLC 経由で登録された CrossRef DOI をもつ文献は、JaLC メタデータに収録されている文献の集合と CrossRef メタデータに収録されている文献の集合の積集合から特定できる。CrossRef メタデータのデータセットを利用すれば積集合を算出することができるが、当該データセットへのアクセスは有料である^{*14}。よって、本稿では CrossRef メタデータの代替として unpaywall データセットを利用する。具体的には、2018 年 9 月 24 日版の unpaywall データセット^{*15}を利用する。unpaywall はあらゆる文献のウェブ上で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であり、Impactstory から提供されている。拡張機能で利用されているデータベースのスナップショットはデータセットとして公開されており、CrossRef DOI を保有する全文の DOI ならびにオープンアクセスに関する情報等を収録している。本稿で使用したデータセットには、99,940,229 件の文献が収録されていた。計算の結果、JaLC メタデータのデータセットに収録されている 6,370,356 件の文献のうち、2,049,903 件 (32.18%) が CrossRef DOI を保有することが判明した。

表 2 に日本の学術出版物として利用するデータセットの概要を示す。JaLC 経由で登録された CrossRef DOI をもつ文献と判断された 2,049,903 件の文献のうち、386,635 件 (18.86%) が COCI で引用文献をオープンにしている。また、unpaywall のデータセットより、2,049,903 件の文献のうち 1,696,647 件 (82.77%) が、ゴールド OA、ハイブリッド OA、グリーン OA などいずれかの方法でオープンアクセスとなっていることが判明した。引用先 (被引用文献)

としての学術出版物のオープン化状況を調査する際には、COCI はあらゆる DOI を被引用文献として収録するため、JaLC メタデータに収録されている全出版物が対象となる。6,370,356 件の JaLC メタデータに収録された文献のうち、544,259 件 (8.54%) が COCI で被引用文献として収録されている。また、714,903 件 (11.22%) が引用文献かつ/あるいは被引用文献として COCI に収録されている。

2.3 学術出版物の分類と収録雑誌

引用データのオープン化の現状を探るにあたり、分類別・収録雑誌別の分析を実施する。そのため、各文献の分類と収録雑誌を特定する必要がある。国立国会図書館が提供する雑誌記事索引採録誌一覧^{*16}を利用することで、文献の ISSN を雑誌に紐付け、収録雑誌を特定する。雑誌記事索引採録誌一覧には 23,910 誌が収録されており、そのうち 17,897 誌の ISSN が記載されている。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^{*17}を利用する。各雑誌には逐次刊行物^{*18}に属する分類が付与されている。

JaLC メタデータに収録されている 6,370,356 件の文献のうち、4,788,310 件 (75.17%) の文献が JaLC メタデータにて ISSN を保有する。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%) が分類を付与されていることが判明した。

本稿では雑誌の内容に関する ZA～ZS の分類に着目する。表 2 の最右列に ZA～ZS の分類をもつ文献件数を示す。

3. 分析結果

本章では引用データのオープン化の現状の分析結果を報告する。3.1 節では、世界の学術出版物における現状との比較を示す。続いて、オープンアクセス状況 (3.2 節)、分

^{*13} JaLC は、JaLC DOI ならびに CrossRef DOI を登録することが可能である。よって、JaLC メタデータは CrossRef DOI をもつ文献も収録されている。

^{*14} <https://www.crossref.org/services/metadata-delivery/plus-service/>, 最終アクセス日: 2019 年 4 月 5 日

^{*15} <https://unpaywall.org/products/snapshot>, 最終アクセス日: 2019 年 4 月 5 日

^{*16} <https://www.ndl.go.jp/jp/data/sakuin/sairokushi.tsv>, 最終アクセス日: 2019 年 4 月 1 日

^{*17} https://www.ndl.go.jp/jp/data/catstandards/classification_subject/ndlc.html, 最終アクセス日: 2019 年 4 月 1 日

^{*18} <http://warp.da.ndl.go.jp/info:ndljp/pid/9484238/www.ndl.go.jp/jp/library/data/z.pdf>, 最終アクセス日: 2019 年 4 月 9 日

表 3 日本の学術出版物と世界の学術出版物における引用データのオープン化の現状

	日本の学術出版物		世界の学術出版物	
	件数	割合 (%)	件数	割合 (%)
引用元として COCI に収録されている文献	386,635	18.86 ¹	24,182,977	24.20 ³
引用先として COCI に収録されている文献	544,259	8.54 ²	38,481,195	21.99 ⁴
COCI に収録されている文献	714,903	11.22 ²	46,530,436	26.59 ⁴

¹ 分母として JaLC メタデータかつ unpaywall に収録されている文献 (≈JaLC 経由で登録された CrossRef DOI をもつ文献) 件数 2,049,903 を利用している。

² 分母として JaLC メタデータに収録されている文献件数 6,370,356 を利用している。

³ 分母として unpaywall に収録されている文献件数 (≈CrossRef DOI をもつ文献件数) 99,940,229 を利用している。

⁴ 分母として DOI の累計発行件数 175,000,000 (概数)^{*19}を利用している。

表 4 オープンアクセス状況別の引用データのオープン化の状況

	全体件数	引用元としてオープン		引用先としてオープン	
		件数	割合 (%)	件数	割合 (%)
オープンアクセスである文献	1,696,647	316,812	18.67	475,955	28.05
オープンアクセスでない文献	353,256	69,822	19.77	63,635	18.01

類 (3.3 節), 収録雑誌 (3.4 節), 出版年 (3.5 節) といった観点から, オープン化の現状について報告する。3.6 節では, 紀要に焦点を当てた分析結果について述べる。

3.1 世界の学術出版物における引用データのオープン化の現状と比較

本節では, 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を, 世界の学術出版物における現状と比較する。

表 3 に, 日本の学術出版物・世界の学術出版物それぞれについて, 引用データがオープンになっている文献件数と割合を示す。世界の学術出版物と比較すると, いずれにおいても日本の学術出版物における引用データのオープン化の割合は低い。特に日本の学術出版物は引用先としての収録が少ないことが明らかである。日本の学術出版物の多くは日本語で記述されているため, 国外の出版物から引用される機会が少ないことが理由の一つとして挙げられる。

3.2 オープンアクセス状況

JaLC メタデータと unpaywall に収録されている文献 (≈JaLC 経由で登録された CrossRef DOI をもつ文献) については, unpaywall によって各文献がオープンアクセスとなっているかという情報が与えられている。2.2 節で述べたとおり, それらの文献の 82.77% がオープンアクセスであると判断されている。本節では, オープンアクセス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

ペン化の傾向に差異が存在するか調査する。調査対象となる文献は, 引用元としてのオープンに注目した場合も, 引用先としてのオープンに注目した場合も, JaLC メタデータと unpaywall に収録されている文献である。

分析結果を表 4 に示す。引用元としてオープンになっている割合は, 若干オープンアクセスでない文献が多いものの, 大差はない。対して, 引用先として COCI でオープンになっている割合については, オープンアクセスである文献が高い。これは, 発見可能性の高さ [3], [4] や引用回数の高さ [1], [3], [4] といった先行研究で観察されたオープンアクセス文献の特長についての結果と合致する。

3.3 分類

本節では, 引用データのオープン化の現状を分類別に探る。各文献の分類は 2.3 節に示す方法によって取得した。分類が取得できなかった場合は, 「不明」という分類に割り当てた。なお, 分類が複数存在する文献については, 1 を与えられた分類数で割り, 各分類に割り当てる。

表 5 に分析結果を示す。引用元としてオープンになっていることに着目すると, 特に化学・化学工業でオープン化が進展していることがわかる。26.12% という数値は, 表 3 に示されている世界の学術出版物における COCI に引用元として収録されている文献の割合 24.20% より大きい。その他, 生物学, 科学技術一般など, STM (科学・医学・工学) 分野において, 引用データのオープン化が進展していることがわかる。対して, 歴史・地理, 哲学・宗教, 芸術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。

続いて, 引用先としてオープンになっていることに焦点

^{*19} <http://www.doi.org/factsheets/DOIKeyFacts.html>, 最終更新日: 2018 年 7 月 4 日, 最終アクセス日: 2019 年 4 月 3 日

表 5 分類別の引用データのオープン化の現状

分類	引用元としてオープン			引用先としてオープン		
	対象文献件数 ¹	COCI での収録件数	割合 (%)	対象文献件数 ²	COCI での収録件数	割合 (%)
政治・法律・行政	4,735.00	99.50	2.10	16,556.00	7.00	0.04
経済	10,179.00	1,534.00	15.07	46,252.50	1,063.00	2.30
社会・労働	10,758.50	1,446.00	13.44	51,997.83	2,506.00	4.82
教育	17,631.00	3,234.00	18.34	75,188.33	2,562.50	3.41
歴史・地理	14,740.00	192.00	1.30	45,401.50	1,293.50	2.85
哲学・宗教	15,104.00	1.00	0.01	27,537.50	41.00	0.15
芸術・言語・文学	6,162.00	595.00	9.66	60,831.50	1,115.00	1.83
科学技術	164,957.75	28,172.00	17.08	387,143.75	46,556.25	12.03
建設工学・機械工学 ³	248,957.25	34,152.50	13.72	356,968.75	51,450.25	14.41
化学・化学工業 ⁴	364,817.75	95,301.50	26.12	428,593.25	138,864.75	32.40
生物学・農林水産	192,388.00	35,436.00	18.42	369,915.50	80,393.50	21.73
心理学・医学・薬学 ⁵	407,761.75	68,203.50	16.73	1,170,096.58	109,580.25	9.37
不明	591,711.00	118,268.00	19.99	3,333,873.00	108,826.00	3.26
合計	2,049,903.00	386,635.00	18.86	6,370,356.00	544,259.00	8.54

¹ JaLC メタデータかつ unpaywall に収録されている文献 (≈JaLC 経由で登録された CrossRef DOI をもつ文献) である。

² JaLC メタデータに収録されている文献である。

³ 建設工学・建設業, 機械工学・工業, 運輸工学, 電気工学・電気機械工業, 原子力工学・工業が含まれる。

⁴ 化学・化学工業, 繊維工学・工業, 食品工学・工業, 金属工学・鉱山工学, 印写工学, その他の工学・工業が含まれる。

⁵ 人類学, 心理学, 医学, 薬学が含まれる。

を当てる。引用元として収録されている割合と比較すると割合の数値は総じて低いものの、分類間を比較した場合の結果は概ね同一である。引用元としてのオープンに注目した場合と同様、化学・化学工業に属する文献が被引用文献として最もオープンになっている。これは、化学・化学工業に属する文献が、同一分類の文献を多く引用しているために、被引用文献としての割合も高くなったのであると推測できる。同様の理由で、STM 分野以外の文献においては、引用先としてオープンになっている割合も低く、その値は総じて 5%未満である。これらの分野は STM 分野と比較すると引用されにくい傾向がある [10] ため、割合が一層低くなっているということも理由として挙げられる。

3.4 収録雑誌

文献のメタデータの付与・登録方法は出版者や編集者によって定められており、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。よって、引用データのオープン化の状況は、収録雑誌によって大きく異なると考えられる。本節では、収録雑誌別に引用データのオープン化の状況を、引用元としてのオープンに注目して分析する。

JaLC メタデータに収録されている文献 6,370,356 件のうち、4,788,310 件の文献に ISSN が付与されている。

4,788,310 件の文献が収録されている雑誌の異なり数は 2,707 誌である。そのうち、unpaywall にも収録されている文献 (≈JaLC 経由で登録された CrossRef DOI をもつ文献) は 2,049,736 件であり、雑誌の異なり数は 1,235 誌である。1,235 誌のうち、857 誌で 1 件以上の文献の引用文献をオープンにしていることが判明した。各雑誌ごとに、COCI に引用元として収録されている文献の割合を算出した。結果の分布を図 1 に示す。このことから、ほとんどの雑誌で、文献の引用文献をオープンにしている割合は 10%以下であることがわかる。収録文献が 100 件以上ある雑誌の中で、引用文献をオープンにしている割合が最も高い雑誌を表 6 に示す。全ての雑誌が STM 分野に属する雑誌であり、いずれも英文誌である。これらの雑誌の多くは Web of Science Core Collection に収録されていてインパクトファクターが付与されていることから、国際的に認知されていることがわかる。

日本の学術出版物の多くは、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。対して、これらの引用文献をオープンにしている割合が高い雑誌は、10 誌中 7 誌が出版プラットフォームとして Nature, Wiley Online Library, Atypen 社の出版プラットフォーム, ScienceDirect といった商業出版社のプラットフォームを利用している*20。これらの商業出版社の慣習が、引用文

*20 J-STAGE も併用されている場合もあるが、最新号の公開は商業出版社のプラットフォームで実施している。J-STAGE ではパッ

献の登録・公開に影響を与えていることが考えられる。

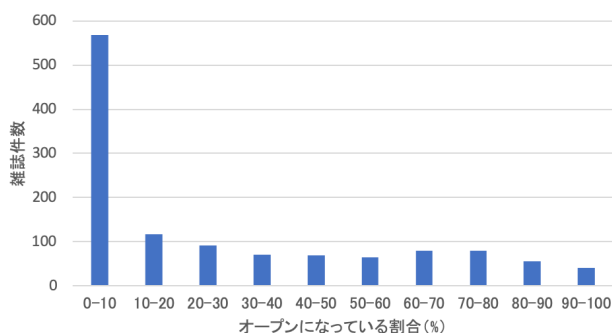


図 1 雑誌別の引用文献をオープンにしている割合の分布

表 6 引用文献をオープンにしている割合が最も高い雑誌

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

3.5 出版年

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、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。本節では、文献の出版年別に引用データのオープン化の状況を調査する。各文献の出版年は JaLC メタデータより取得した。JaLC メタデータでは、収録されている文献 6,370,356 件のうち、3 件を除く 6,370,353 件の文献に出版年が与えられている。

図 2 に、1991～2018 年の期間において JaLC メタデータと unpaywall に収録されている文献（～JaLC 経由で登録された CrossRef DOI をもつ文献）がどの程度引用文献をオープンにしているかについて文献の出版年ごとに集計した結果を示す。なお、本稿で利用しているデータセットは 2018 年は 9 月に公開されたものであるため、2018 年に出版された文献は他の年と比較すると少ない。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 COCI でオープンになっていることがわかる。対して、過去の文献では、オープン化は進んでいない。日

クナンバーを公開している。

本の状況を世界の状況と比較するために、図 3 に、1991～2018 年の期間において unpaywall に収録されている文献（～CrossRef DOI をもつ文献）がどの程度引用文献をオープンにしているかについて文献の出版年ごとに集計した結果を示す。日本の状況と比較すると、出版年がオープン化の割合に与える影響は少ない。2000 年以前は、日本の学術出版物のオープン化の割合より高い数値である 18%前後で推移している。その後、2003 年以降は、2013 年と 2017 年を除く出版年で、日本の学術出版物のオープン化の割合は世界の学術出版物と比較して高い。よって、近年に注目すると、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。新しい文献と比較すると過去の文献が閲覧される回数は少ない [1] が、日本の学術出版物においては過去の文献の引用文献をオープンにしていくことが課題となるだろう。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の一つとして、大手商業出版社による影響が挙げられる。Heibi らの調査 [11] では、Elsevier が出版元である CrossRef DOI が付与された 16,773,716 件の文献のうち、11,020,314 件が引用文献をメタデータに記述している。しかし、それら全件はオープンデータとして公開されていない。それらを公開することで、各出版年におけるオープン化の割合は 10%以上増加する。

図 4 に、1991～2018 年の期間において JaLC メタデータに収録されている文献が被引用文献としてどの程度オープンになっているかについて文献の出版年別に示す。1991～2010 年の期間に着目すると、いずれの出版年でも文献のうち約 10%程度が被引用文献として COCI に収録されており、出版年による変化は乏しい。2010 年以降の割合が緩やかに減少しているが、これは近年出版された文献は過去に出版された文献と比較すると、被引用の機会が少なかったことが理由として挙げられる。時間が経過するにつれて、これらの出版年の文献の被引用回数は上昇するだろう。図 5 に、1991～2018 年の期間において unpaywall に収録されている文献（～CrossRef DOI をもつ文献）が被引用文献としてどの程度オープンになっているかについて文献の出版年ごとに集計した結果を示す。日本の学術出版物で観察された傾向と同様、2009 年頃までは一定の割合で推移しているものの、その後に割合は減少する。ただし日本の学術出版物と比較すると、オープン化の割合は常に高い。これは、多くの場合英語で記載されている世界の学術出版物は対象とする読者が多く引用されやすいということが理由として挙げられる。

3.6 紀要

大学などの研究機関が発行する学術雑誌である紀要は、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 [12]。出版者のプラットフォームでは各文

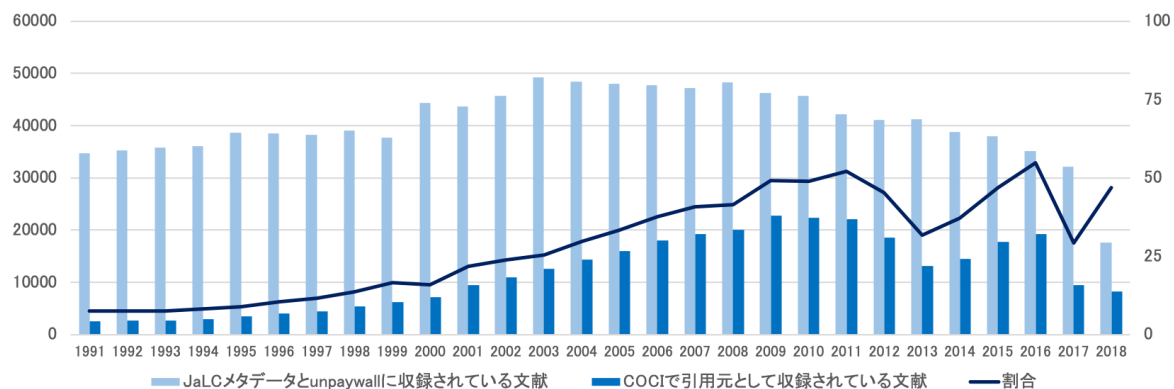


図 2 日本の学術出版物における出版年ごとの引用元としてのオープン化の状況

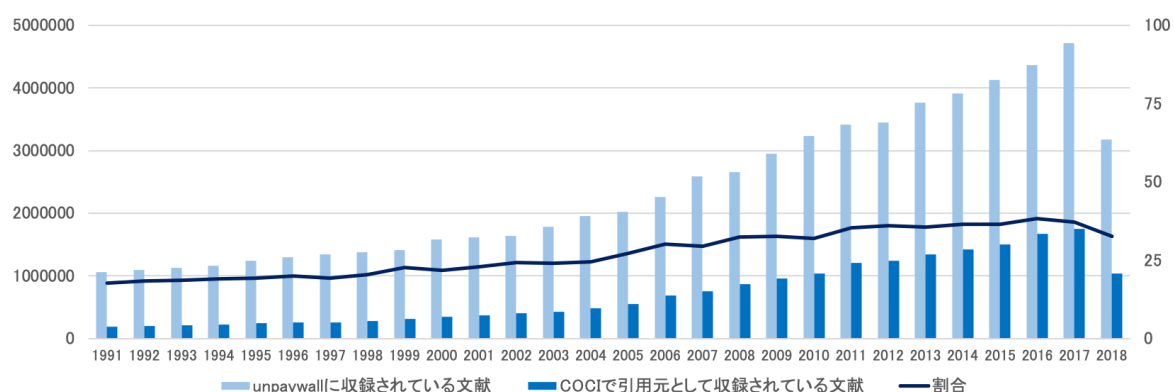


図 3 世界の学術出版物における出版年ごとの引用元としてのオープン化の状況

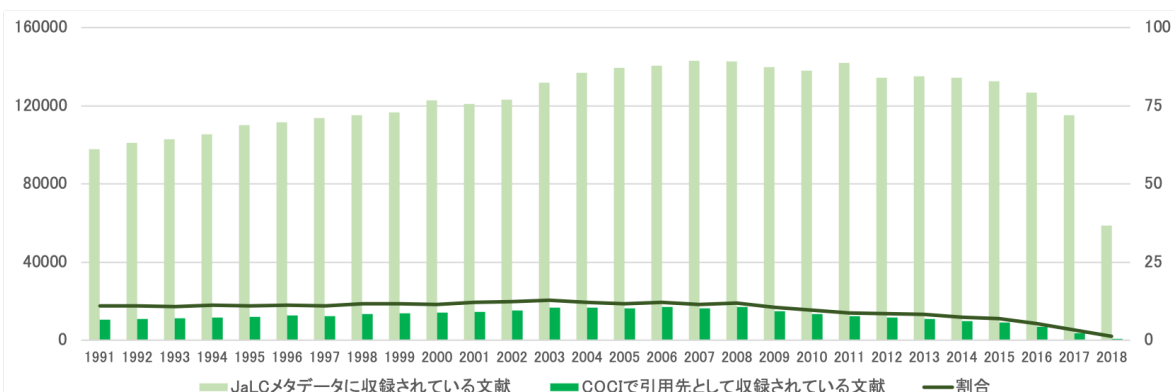


図 4 日本の学術出版物における出版年ごとの引用先としてのオープン化の状況

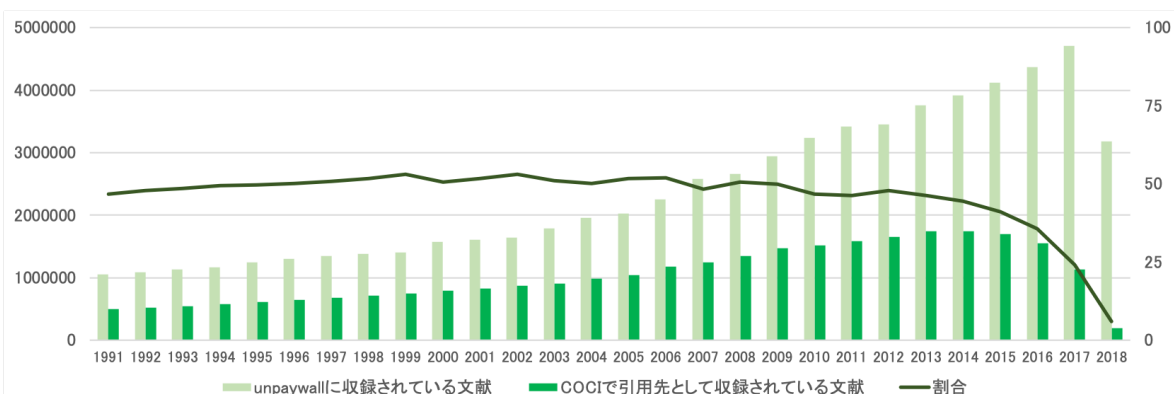


図 5 世界の学術出版物における出版年ごとの引用先としてのオープン化の状況

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場合が多く、文献に述べられている科学的知見の検証可能性の向上に寄与している。対して、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場合が多く、同様の機能が望まれる。このような機能を検討するにあたっては文献の引用データがオープンになっていることが望ましいことから、本節では紀要に収録されている文献に焦点を当てて、引用データのオープン化の現状を分析する。

紀要に収録されている文献として、国立国会図書館分類ZV1(紀要)をもつ文献18,079件を抽出した。雑誌記事索引採録誌一覧では、掲載されている23,910誌のうち3,029誌にZV1の分類が与えられている。抽出した18,079件の文献は、47誌の紀要のうちいずれかに収録されていることが判明した。雑誌記事索引採録誌一覧では、紀要に分類される3,029誌のうち2,730誌にISSNが付与されている。JaLCメタデータにて紀要の文献のISSNが収録されていないため紐付けができず、分析対象が47誌と少数になってしまったと考える。

18,079件のうち、6,533件がunpaywallにも収録されている文献(≈JaLC経由で登録されたCrossRef DOIをもつ文献)である。6,533件の文献は、日本學士院紀要、Proceedings of the Japan Academy, Synthesiology, 社会技術研究論文集, 人間環境学研究のいずれかに収録されている。そのうち2,131件(32.62%)が引用元としてCOCIに収録されている。前述の5誌にいずれにおいても、引用データをオープンにしていることが判明した。32.62%という数値は、表3にある日本の学術出版物、世界の学術出版物についての数値と比較しても高いといえるが、標本数が少ないため「紀要において引用データのオープン化が進展している」といった結論を導き出すことはできない。

引用先としてのオープンに注目すると、18,079件の文献のうち2,131件(11.79%)が引用先としてCOCIに収録されていることが判明した。

4. おわりに

本稿では、I4OCの活動を受けて公開されたCOCIの引用データを利用することで、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を探った。世界の学術出版物の引用文献のオープン化の割合が24.20%であるのに対して日本の学術出版物においては18.86%であり、海外と比較すると引用データのオープン化は進んでいない。しかし過去10年に注目すると、日本の学術出版物の引用文献のオープン化の割合は海外と比較して高い。また、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ことが判明した。対して、人文学系分野オープン化の割合は総じて低い。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

が重要となるだろう。

また、J-STAGEでは各文献の引用文献・被引用文献を表示する機能は実装されているが、多くの機関リポジトリにはこのような機能は実装されていない。ユーザの情報検索行動に関する研究[7], [8]ではこれらの機能が重要であることが示されている。引用データを利活用して学術の発展の系統を辿れるようにするためにも、これらの機能の実装が望まれるだろう。

謝辞 本研究の一部は、京都大学東南アジア地域研究研究所・東南アジア研究の国際共同研究「東南アジア地域研究資料のオープン・サイエンス化に向けたとりくみ」、JSPS科研費若手研究18K13235の助成を受けたものです。

参考文献

- [1] Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J. and Haustein, S.: The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles, *PeerJ*, Vol. 6, p. e4375 (2018).
- [2] Smith, L. C.: Citation analysis, *Library Trends*, Vol. 30, No. 1, pp. 83–106 (1981).
- [3] Wang, X., Liu, C., Mao, W. and Fang, Z.: The open access advantage considering citation, article usage and social media attention, *Scientometrics*, Vol. 103, No. 2, pp. 555–564 (2015).
- [4] Eysenbach, G.: The open access advantage, *Journal of Medical Internet Research*, Vol. 8, No. 2 (2006).
- [5] Lauscher, A., Eckert, K., Galke, L., Scherp, A., Rizvi, S. T. R., Ahmed, S., Dengel, A., Zumstein, P. and Klein, A.: Linked Open Citation Database: enabling libraries to contribute to an open and interconnected citation graph, *ACM/IEEE on Joint Conference on Digital Libraries*, ACM, pp. 109–118 (2018).
- [6] Peroni, S., Shotton, D. and Vitali, F.: One year of the OpenCitations Corpus, *International Semantic Web Conference*, Springer, pp. 184–192 (2017).
- [7] Kacem, A. and Mayr, P.: Analysis of search stratagem utilisation, *Scientometrics*, Vol. 116, No. 2, pp. 1383–1400 (2018).
- [8] Kacem, A. and Mayr, P.: Analysis of footnote chasing and citation searching in an academic search engine, *arXiv preprint arXiv:1707.02494* (2017).
- [9] Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-j. P. and Wang, K.: An overview of Microsoft Academic Service (MAS) and applications, *International Conference on World Wide Web*, ACM, pp. 243–246 (2015).
- [10] 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).
- [11] Heibi, I., Peroni, S. and Shotton, D.: Crowdsourcing open citations with CROCI-An analysis of the current status of open citations, and a proposal, *arXiv preprint arXiv:1902.02534* (2019).
- [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).