

# 学術論文違法共有サイトへの日本からのアクセス状況 —Sci-Hub の公開ログ・データを元にして—

栗山正光

首都大学東京学術情報基盤センター

mtkuri@tmu.ac.jp

## 抄録

Sci-Hub という学術論文違法共有サイトが議論を呼んでいるが、その利用者は世界中に広がっており、日本からも多くのアクセスがあることが明らかになっている。本研究では公開されている Sci-Hub のログ・データの中から、日本からのアクセス・データを抽出し、分析した。その結果、地理的には東京からのアクセスが 6 割を占め、一極集中の様相を呈していること、また、アクセスが集中する特定の日が見られること、ダウンロードが多い出版社などは世界的な傾向と一致することなどが判明した。

## 1. 研究の背景、目的

学術論文はいわゆる電子ジャーナルに掲載され、インターネットを介してアクセスされるのが常態となっている。その一方で電子ジャーナルの価格は上昇を続け、大学図書館が購読を維持できず、多くの研究者が必要な論文を入手できないという事態にも陥っている。

これに対して、誰でも無料で学術論文を読めるようにしようというオープンアクセス(OA)運動が起こっている。OA は著作権を尊重しつつ行われるもので、著者が出版社の同意を得て論文原稿をリポジトリ等で公開するグリーン OA、出版費用を著者の側で負担する(研究助成金等で賄う場合を含む)ゴールド OA といった方法がある。しかし、どちらも普及にはほど遠く、まだまだ購読料を払った人にしかアクセスできない論文が多い。

こうした状況の中、カザフスタンの大学院生アレクサンドラ・エルバキヤン(Alexandra Elbakyan)が立ち上げた Sci-Hub という学術論文共有サイトが議論を呼んでいる。ここでは有料の電子ジャーナル掲載論文を、支持者から提供された、あるいは不正に入手したパスマ

ードで取得し、大量に蓄積して無料提供している。明らかに違法で、大手出版社から訴えられたりもしているが、ドメイン名を変えつつ存在し続けている。

ジョン・ボハノン(John Bohannon)は『サイエンス』誌にこの Sci-Hub に関する記事を発表し、先進国も含めた世界中にこのサイトの利用者が広がっていることを明らかにした<sup>1)</sup>。彼はエルバキヤンと接触し、Sci-Hub サーバーのログ・データを譲り受け、それに基づいて利用の傾向を分析しているのだが、そのログ・データは記事に付随するデータとして公開されている<sup>2)</sup>。

本研究は、その公開ログ・データを元にして、Sci-Hub への日本からのアクセス状況を調べたものである。研究の目的は、日本からはこうした違法サイトがどの程度利用されているのか、また利用の傾向はどのようなものかを、ボハノンが示した世界的な状況と比較しつつ、明らかにすることである。

## 2. 研究の方法

上記公開ログ・データは、2015年9月1日

から翌年2月28日までの6か月間にわたる約2,800万件のダウンロード要求の記録である。ただし、11月5日から21日までのデータは欠落している。データ項目としては、日時、論文のDOI、IPアドレスから割り出された地理的情報（国名、都市名、経緯度）がある（プライバシー保護のため元のIPアドレスはわからないようにしてある）。

このデータをパソコンにダウンロードし、日本からのアクセス・データを抽出した。この作業にはデータベース・ソフト（マイクロソフト・アクセス）を利用した。その後、抽出されたデータを表計算ソフト（エクセル）に読み込んで加工し、要素別に集計して分析した。最初からエクセルに読み込まなかったのは、データ量が多すぎて処理できなかったからである。

### 3 地理的要素による集計結果

全レコードを国・地域別に集計すると、223の国・地域に分けられる。不明(N/A)のものもあり、全件数は27,257,271に減る。表1はそのうち上位30か国を示したものである。中国が1位で440万件以上、2位インドが340万件、3位イランが260万件と、この3か国が圧倒的に多く、ほぼ4割を占める。

注目すべきは、電子ジャーナルへのアクセス環境が整っていると思われる米国が115万件で5位となっていることである。フランス、ドイツも15位以内に入っている。すなわち、貧しい発展途上国の研究者が主に利用しているというわけでは決してないのである。わが日本も、全体の1%に満たないとは言え、約23万5千件で28位となっている。

日本のみを抽出したデータについて、都市別に集計した結果、384の市（東京は区）・郡からアクセスがあったことが判明した（“N/A”になっていた20件は経緯度が大阪府庁のものだったため大阪市にカウントした）。都市名の元

表1 国別ダウンロード数

	国・地域	件数	%
1	China	4,456,071	16.35
2	India	3,414,581	12.53
3	Iran	2,631,035	9.65
4	Russia	1,521,434	5.58
5	United States	1,150,961	4.22
6	Brazil	1,021,540	3.75
7	Egypt	801,105	2.94
8	Indonesia	780,873	2.86
9	France	511,145	1.88
10	Mexico	503,093	1.85
11	Morocco	497,306	1.82
12	Pakistan	493,996	1.81
13	Germany	462,011	1.70
14	Portugal	434,721	1.59
15	Spain	426,412	1.56
16	Chile	425,596	1.56
17	Colombia	414,783	1.52
18	Argentina	410,986	1.51
19	Turkey	393,308	1.44
20	Malaysia	373,752	1.37
21	Peru	373,325	1.37
22	Vietnam	369,246	1.35
23	Ukraine	367,047	1.35
24	Tunisia	339,592	1.25
25	Algeria	295,478	1.08
26	Italy	295,015	1.08
27	Greece	285,452	1.05
28	Japan	234,755	0.86
29	United Kingdom	218,631	0.80
30	South Korea	177,023	0.65

データはローマ字だが、漢字に置き換えた。表2に上位20都市のアクセス（ダウンロード）件数と全件数に対する割合を示す。東京の新宿区と千代田区の数が突出して多く、名古屋、京都、つくばといった大学・研究所が多い都市がそれに続いている。

人口の多い大都市が多いのは当然と思われるので、表3に人口比（人口一人当たりの件数）による上位20件を示した。人口は2016年6月1日現在の推計人口<sup>3)</sup>である。ただし、新宿区と千代田区は常住人口だと数字が大きく出過ぎるので昼間人口<sup>4)</sup>とした。しかしそれでもやはり、新宿区、千代田区が1,2位を占め、研究学園都市を有するつくば市が3位となる。

表 2 都市別ダウンロード数

	都市	都道府県	件数	%
1	新宿区	東京都	93,077	39.65
2	千代田区	東京都	39,935	17.01
3	名古屋市	愛知県	7,710	3.28
4	京都市	京都府	7,013	2.99
5	つくば市	茨城県	6,171	2.63
6	大田区	東京都	4,737	2.02
7	大阪市	大阪府	4,383	1.87
8	富山市	富山県	3,922	1.67
9	千葉市	千葉県	3,377	1.44
10	福岡市	福岡県	3,305	1.41
11	横浜市	神奈川県	2,988	1.27
12	川崎市	神奈川県	2,779	1.18
13	岡山市	岡山県	2,468	1.05
14	函館市	北海道	2,459	1.05
15	相模原市	神奈川県	2,343	1.00
16	長崎市	長崎県	2,119	0.90
17	宇都宮市	栃木県	2,014	0.86
18	仙台市	宮城県	1,745	0.74
19	浜松市	静岡県	1,690	0.72
20	札幌市	北海道	1,609	0.69

表 3 都市別ダウンロード数（人口比順）

\* 新宿区、千代田区の人口は昼間人口

	都市	都道府県	件数	人口	人口比
1	新宿区	東京都	93,077	737,252	0.1262
2	千代田区	東京都	39,935	812,360	0.0492
3	つくば市	茨城県	6,171	229,451	0.0269
4	摂津市	大阪府	1,011	85,016	0.0119
5	鹿沼市	栃木県	1,156	97,916	0.0118
6	富山市	富山県	3,922	418,529	0.0094
7	函館市	北海道	2,459	266,785	0.0092
8	宇部市	山口県	1,459	168,511	0.0087
9	由利本荘市	秋田県	627	78,763	0.0080
10	高座郡	神奈川県	357	48,046	0.0074
11	大田区	東京都	4,737	721,978	0.0066
12	観音寺市	香川県	387	59,073	0.0066
13	長崎市	長崎県	2,119	430,347	0.0049
14	京都市	京都府	7,013	1,475,042	0.0048
15	知多郡	愛知県	728	161,575	0.0045
16	宇都宮市	栃木県	2,014	519,217	0.0039
17	鳥取市	鳥取県	728	192,912	0.0038
18	千葉市	千葉県	3,377	974,335	0.0035
19	岡山市	岡山県	2,468	720,745	0.0034
20	名古屋市	愛知県	7,710	2,302,696	0.0033

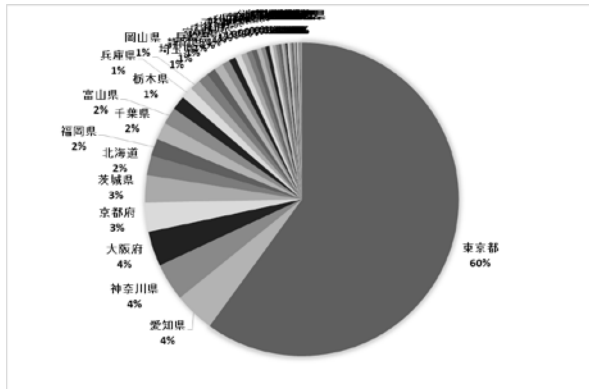


図 1 都道府県別ダウンロード数 (%)

摂津、鹿沼といった地方都市がここで浮上してくるが、このあたりからは低い数値で差もあまりないため、意味がない順位かもしれない。

さらに、都市が存する都道府県別に件数を集計した。その結果を円グラフで示したのが図 1 である。東京都が 6 割を占め、あとは皆 4%以下である。こんなところにも東京への一極集中が見られる。

#### 4 時間的要素による集計結果

図 2 は件数を日毎に集計した結果を時系列で示したグラフである。前後の日比べて突出して多い日が 3 日ほど見られる。中でも 2015 年 12 月 12 日は、この日だけで 11,939 件もの

ダウンロードがあった特異日である。このうち、10,655 件は 2 時 50 分から 18 時 1 分にかけて新宿区から継続的にリクエストされた米土木学会(American Society of Civil Engineers)の論文である。テキストマイニングか何かのため、特定の人物が機械的に大量ダウンロードを行ったと推測される。

『サイエンス』誌に掲載された世界全体のグラフでは、凹凸は当然あるものの、こうした極端な突出は見られない。日本の場合、データ数が世界に比べれば圧倒的に少ないため、外れ値の影響が目立つと言えよう。

#### 5 出版社・論文による集計結果

DOI のプレフィックスを元に出版社別に集計してみると、日本からは 630 の出版社の論文がダウンロードされていることがわかった。上位 15 社で 74%、30 社で 84%を占めており、偏りの激しいロングテールになっている。特に、首位のエルゼビアは 1 社で 25%と群を抜いているが、同社は世界では 34%とさらに大きなシェア (?) を占めている。ワイリー・ブラックウェルのように一つの会社が複数のブラン



図2 日毎のダウンロード数

ドに分けられている場合もあるので、実際の集中度はさらに高くなる。

図3は世界でダウンロード数の多い出版社上位15社を示したものである。横軸は世界でのダウンロード数、縦軸は日本でのダウンロード数である。一見してわかるとおり、両者には強い相関関係がある。すなわち、世界でも日本でも同じような出版社の論文が多くダウンロードされている。15社中13社が日本の上位15社に含まれており、ベスト3はどちらもエルゼビア、シュプリンガー、IEEEである。

これに対し、多くダウンロードされた論文は日本と世界で異なっている。『サイエンス』誌に発表されている世界上位10本の論文は、日本では1本が4回、残りはすべて1, 2回ダウンロードされているに過ぎない。

出版社別の件数が一部の大出版社に大きく偏っているのに対し、論文別の件数にはあまり偏りはない。日本では234,755回で207,095本の論文がダウンロードされており、1回しかダウンロードされていない論文が8割を占める。

日本で一番多くダウンロードされている論文は51回、世界では7,988回だが、いずれも全ダウンロード数の0.02%程度に過ぎない。当然のことながら、研究者が必要とする論文は多種多様であり、それがこのSci-Hubの利用データにも表れていると言えよう。

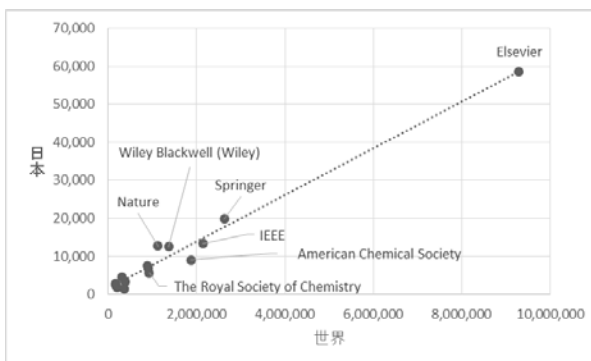


図3 ダウンロードの多い出版社

#### 参考文献

- (1) Bohannon, J. “Who’s downloading pirated papers? Everyone,” *Science*, 352(6285), 2016, p. 508-512.  
<http://dx.doi.org/10.1126/science.352.6285.508>, (参照 2016-10-07).
- (2) Elbakyan, A., Bohannon, J. *Data from: Who’s downloading pirated papers? Everyone*. Dryad Digital Repository, 2016.  
<http://dx.doi.org/10.5061/dryad.q447c>, (参照 2016-10-07).
- (3) 「ウィキペディア. 日本の市の人口順位」,  
<https://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E5%B8%82%E3%81%AE%E4%BA%BA%E5%8F%A3%E9%A0%86%E4%BD%8D>, (参照 2016-10-10).
- (4) 「東京都昼間人口の予測—統計データ—平成27年3月」<http://www.toukei.metro.tokyo.jp/tyosoku/ty-data.htm>, (参照 2016-10-10).