

# 研究データ公開に関する学術雑誌のポリシー分析

池内有為(筑波大学大学院図書館情報メディア研究科)

oui@slis.tsukuba.ac.jp

## 1. はじめに

### 1.1 研究背景

近年、研究成果として論文などの出版物だけではなく、研究に用いたデータをも公開し、共有する動きが自然科学分野を中心に社会科学分野に至るまで拡大しつつある。その背景には、他の研究者がデータを再利用することによる研究の効率化、研究結果の検証を可能にすること、そして公的資金による研究成果を市民に公開することに対するニーズがある。また、学術雑誌や研究助成機関によるデータ公開の義務化によっても公開が推進されていることが報告されている。データ公開に関する学術雑誌のポリシーは、データ共有が盛んに行われている生命科学分野を中心として調査されている。しかし、自然科学分野から社会科学分野全体の現在の動向を捉えるまでには至っていない。そこで本研究では18分野を対象とした調査を行い、分野ごとの特徴を比較するとともに、雑誌のインパクトファクター(IF)やオープンアクセス(OA)等の要因との関係について分析を行った。

### 1.2 先行研究

2007年にPiwowarら<sup>1)</sup>は、マイクロアレイによる遺伝子発現解析に関する70誌のデータ

公開に関するポリシーを調査している。その結果、42誌(60%)がデータ公開ポリシーを掲げており、さらに学協会の方が商業出版社よりも多いこと、IFの高い雑誌の方が多いこと、OAの4誌はすべて何らかのポリシーを記載していることが明らかになった。Alsheikh-Aliら<sup>2)</sup>は、2009年にトムソン・ロイター社のJournal Citation Report(JCR)を用いてIFの上位50誌を対象とした調査を行い、44誌(88%)が何らかのポリシーを記述していることを明らかにした。ただし、IFの高い雑誌はPiwowarら<sup>1)</sup>の研究と同様に生命科学分野が中心であり、ポリシーもマイクロアレイや核酸配列データのリポジトリ登録に関する内容であった。

Vlaeminck<sup>3)</sup>が2013年に発表した論文によれば、ドイツ、オーストリア、スイスで出版された主要な経済学分野の雑誌141誌のうち、29誌(21%)にデータ可用性(availability)ポリシーが示されている。また、Stoddenら<sup>4)</sup>は確率統計、数理・計算生物学等の計算科学分野170誌のデータ、コード、補足資料の公開に関するポリシーを2011、2012年に調査している。その結果、一年間で公開ポリシーを持つ雑誌が56誌(33%)から64誌(38%)に増加していることや、IFが高いほど公開ポリシーを持つ割合の高いことが明らかになった。

表 1. 学術雑誌のデータ公開ポリシーに関する先行調査の概要

著者	調査年	調査対象分野	雑誌数	掲載誌数
Piwowarら	2007	生命科学分野	70	42 (60%)
Alsheikh-Aliら	2009	生命科学分野等 (IFの高い50誌)	50	44 (88%)
Vlaeminck	(2013)	経済学(ドイツ, オーストリア, スイス)	141	29 (21%)
Stoddenら	2011	計算科学	170	56 (33%)
	2012		170	64 (38%)

## 2. 研究方法

### 2.1 調査の概要

本研究における「(研究)データ」とは、研究のために生成したが論文には掲載されていない生データとし、画像、動画、テキスト、コード、データセット等、データの形式やサイズは問わないものとする。また、その「公開」とは、GenBank などのパブリックリポジトリや、出版者、著者のサーバに登録されていることとする。

本研究が対象とする「データ公開ポリシー」とは、各雑誌の投稿規定や執筆要綱に掲載されているデータ公開に関する記述であり、各雑誌のウェブサイトで行った。

### 2.2 分野と雑誌の選定

複数の先行研究<sup>14)</sup>によって IF の高い雑誌ほどデータ公開ポリシーの掲載率が高いことが示されている。そこで JCR(2012)を用いて対象とする分野の IF の高い学術雑誌を上位 10 誌ずつ抽出して調査対象とすることとした。

自然科学分野に関しては、JCR(2012)の Science Citation Index (SCI) の 176 カテゴリを米国科学審議会 (National Science Board) の科学工業指標 (Science and Engineering Indicators) に用いられている 11 の区分<sup>5)</sup>に分類した。さらにデータ公開が進んでいるとされている Biological Sciences と Medical Sciences については、下位区分を設けて細分し、分野内での比較を可能にした。ただし、科学工業指標の下位区分はそれぞれ 12, 34 区分あり、詳細になりすぎる。また、JCR のカテゴリは同一の雑誌に複数付与されている。そこで、まず、Biological Science に相当する JCR の 22 カテゴリと Medical Sciences に相当する 43 カテゴリをそれぞれ統合して、2 分野の IF 上位 500 誌を抽出した。合計 1,000 誌分のデュエイ番号をシリアルズ・ソリューション社の Ulrichsweb.com から取得し、これを基に下位区分として分類した。以上の手続きによって得

られた 30 分野から、生物学、生理学、生化学、遺伝学、生態学、微生物学、動物学、医学、薬理学、疾患学、外科学、天文学、地球科学、物理学、化学、計算機科学の 16 分野を調査対象として選定した。

社会科学分野については、Social Science Citation Index (SSCI) の 56 カテゴリから、先行研究で調査が行われた経済学と、公開された社会調査データや統計データの再分析が行われている社会学を対象とした。

以上、自然科学 16 分野と社会科学 2 分野、計 18 分野の IF の高い査読誌 10 誌について、データ公開ポリシーを調査した。なお、レビュー論文はその性質上、データを公開しているとは考え難いため、レビュー誌は除外した。レビュー誌かどうかの判断は、雑誌の編集方針や投稿規定、掲載論文を確認することによって行った。また、速報、短報誌についても同様に判断して対象外とした。

### 2.3 調査内容

Stodden<sup>4)</sup>の調査を参考に、データ公開ポリシーを、著者への要求の度合いに応じて 5 段階に分類して調査した(表 2)。

表 2. データ公開ポリシーの分類

pt	著者への要求度合	表現
4	要求(査読・出版条件)	required, must,
3	要求	should
2	奨励	encouraged
1	言及のみ	(implied)
0	記載なし	(no mention)

査読や出版の条件(condition)としてデータを公開するよう義務化している場合を 4 ポイント(pt)として、以下、要求しているが査読や出版の条件にはしていない場合を 3pt、公開を奨励している場合を 2pt、データ公開について

何らかの言及があるものの、特に著者に公開を求めている場合を 1pt、全く記載がない場合を 0pt とした。

## 2.4 補足資料によるデータ公開

Stodden ら<sup>4)</sup>は、データ公開ポリシーとは別に、補足資料(Supplemental Materials)のポリシーを調査している。予備調査の結果、補足資料とは、印刷版の本文には含められないが、オンライン版で提供されているデジタルデータであった<sup>6)</sup>。その内容は、論文に関連する図、表、画像、動画、大規模データセット等であり、本文にリンクして参照される場合と、別途付録として公開される場合がある。また、査読対象として提出が義務づけられている場合もあれば、任意の場合もある。

これらのデータは、他の研究者が再利用したり、研究結果の検証を行ったりすることが可能であり、本研究で対象としている「データ」と本質的には同一のものである。その登録先が外部のリポジトリか出版者等のサーバかという違いでしかない。ただし、それが「公開」されているかどうかは確実ではない。つまり、掲載誌や論文が OA の場合のみ「公開」されていると考えられる。しかし例外的に、有料論文であっても補足資料が公開されている場合もある。

一方で、データ公開ポリシーはないものの、補足資料としてデータを公開するよう要求する雑誌も見られた。そこで、これらの論点をふまえた上で、補足資料もデータ公開の一形態とみなし、あわせて調査を行った。

## 3. 結果と考察

### 3.1 分野別研究データ公開ポリシー

分野別データ公開(Data Sharing: DS)ポリシーの調査結果、および IF と OA 誌のタイトル数を表 3 に示す。DS 列には、表 2 に示した 5 段階の加重平均値を示している。IF は各分野の 10 誌の平均値である。DS と IF の無相関

検定を行ったところ、相関係数は 0.529 であり、有意水準 5% で有意であった。外科学、社会学は、全ての雑誌でデータ公開ポリシーが見られなかったものの、補足資料については、それぞれ 4 誌、2 誌で公開可能であった。

表 3. 分野別研究データ公開ポリシー

分野	DS	IF	OA
微生物学	9.25	7.14	3
生理学	8.50	13.85	0
生化学	8.50	8.87	1
医学	8.25	22.46	1
生物学	8.25	9.62	3
遺伝学	8.00	9.70	1
化学	6.25	18.68	0
地球科学	6.25	9.19	2
物理学	5.75	18.73	0
生態学	5.75	8.25	0
疾患学	5.50	18.46	0
動物学	5.25	3.41	1
薬理学	2.50	6.61	0
経済学	2.25	4.16	0
天文学	2.00	6.41	0
計算機科学	0.75	5.29	1
外科学	0	5.09	0
社会学	0	3.03	0

微生物学は 10 誌ともデータ公開を要求しており、うち 7 誌は査読・出版の要件としている。Biological Sciences は生理学、生化学、生物学、遺伝学等も同様にデータ公開を要求する雑誌が多く、補足資料においてもデータの公開が要求または推奨されている。例示されているリポジトリも GenBank, EMBL, DDBJ 等に集中しており、分野としてのコンセンサスが確立していると考えられる。

Medical Sciences は特にデータの信頼性が求められる分野であり、今後データ公開が進

展することが予想される。英国医学会が発行するBMJはOpen Data Campaignを展開している<sup>7)</sup>が、一方で患者の個人情報保護といった倫理的側面からデータ公開が困難でもある。本調査の結果も分野によって差が見られ、特に外科学はデータ公開ポリシーが10誌とも掲載されており、補足資料も強く要求されていない。

経済学はデータ公開を査読・出版要件とする雑誌はみられなかったものの、10誌中3誌がデータ公開を要求しており、先行研究<sup>3)</sup>のData Availability Policyの掲載率21%とほぼ同様の結果となった。ただし、補足資料については9誌が何らかの言及を行っており、特にデータ設計、被験者、プログラム、生データを全てAppendixや補足資料に掲載するよう求める*American Economic Review*のData Availability Policy<sup>8)</sup>に準拠する雑誌が多くみられた。

### 3.2 OAとデータ公開ポリシー

180誌中13誌がOAであったが、全ての雑誌のデータ公開ポリシーは「要求(査読・出版条件)」もしくは「要求」であった。これは全てのOA誌が何らかの公開ポリシーを持つという先行研究と同様の結果である。ただし、先行研究も本研究も調査タイトル数が少ないため、OA誌であればデータ公開ポリシーがあるとはかならずしも言い切れないであろう。

### 3.3 出版社と雑誌のポリシー

同一の出版社や学協会が提供していても、雑誌ごとにデータ公開ポリシーは異なっていた。雑誌ごとに個別にポリシーを策定している場合と、出版社の掲げる複数のポリシーやガイドラインから雑誌の方針に適合するものを選択する場合とがある。後者の例として、エルゼビア社はData deposit and linkingというガイドラインを掲げている<sup>9)</sup>が、同社が提供している

天文学分野の雑誌である*Icarus*はこれを採用し、*Astroparticle Physics*は採用していない。

### 謝辞

本研究は、2013年度日本図書館情報学会研究助成を受けて実施しています。この場を借りて厚く御礼申し上げます。

### 参考文献

- <sup>1</sup> Piwowar, Heather A; Chapman, Wendy W. A review of journal policies for sharing research data. *Nature Precedings*. 2008, 3p.
- <sup>2</sup> Alsheikh-Ali, Awai A; Qureshi, Waqas; Al-Mallah, Mouaz H; Ioannidis, John P. A. Public Availability of Published Research Data in High-Impact Journals. *PLoS ONE*. 2011, Vol. 6, Issue. 9, e24357. doi:10.1371/journal.pone.0024357
- <sup>3</sup> Vlaeminck, Sven. Data management in scholarly journals and possible roles for libraries: some insights from EDaWaX. *LIBER Quarterly*. 2013, Vol. 23, No. 1, p.48-79.
- <sup>4</sup> Stodden, Victoria; Guo, Peixuan, Guo; Ma, Zhaokun. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journal. *PLOS ONE*. 2013, Vol. 8, Issue 6, e67111.
- <sup>5</sup> National Science Board. "Appendix table 5-24: Fields and subfields of S&E publication data." *Science and Engineering Indicators 2010*. 全13区分からPsychologyとSocial Sciencesを除く11区分を自然科学分野の区分として採用した。 <http://www.nsf.gov/statistics/seind10/append/c5/at05-24.pdf> (accessed 2013-9-1)
- <sup>6</sup> *Journal of Molecular Cell Biology* は、"Data that are integral to the manuscript but impractical to include in the printed journal may be presented as Supplementary Data."としている。"Supplementary Data" in Instructions to Authors. [http://www.oxfordjournals.org/our\\_journals/jmcb/for\\_authors/](http://www.oxfordjournals.org/our_journals/jmcb/for_authors/) (accessed 2013-9-1)
- <sup>7</sup> BMJ. Open Data. <http://www.bmj.com/open-data>
- <sup>8</sup> The American Economic Review. Data Availability Policy. <http://www.aeaweb.org/aer/data.php> (accessed 2013-10-1)
- <sup>9</sup> "Data deposit and linking" in Guide for Authors. リポジトリに生のデータセットを登録することを推奨し、さらにScienceDirectに掲載する論文へのリンクを提供している。Elsevier <http://www.elsevier.com/about/content-innovation/database-linking> (accessed 2013-9-1)