

学術雑誌によるデータ共有ポリシー：分野間比較と特徴分析

池内有為 (筑波大学大学院図書館情報メディア研究科)

逸村裕 (筑波大学図書館情報メディア系)

【抄録】

学術雑誌によるデータ共有ポリシーの分野別状況を明らかにするため、22 分野各 10 誌の投稿規定を調査した。リポジトリにデータを公開して論文に識別子を記すポリシーと補足資料のポリシーを、それぞれの要求の強度に従って 4 段階に分類した。

生物・医学の 10 分野はリポジトリによるデータ共有ポリシーの掲載率や強度が高い傾向にあり、共通のリポジトリを例示していたが、農学、薬理学・毒物学、精神医学・心理学の掲載誌はそれぞれ 7, 6, 2 誌であり領域内で差がみられた。地球科学、宇宙科学、社会科学はデータ共有が盛んであるが、リポジトリによるポリシーの掲載誌は 7, 6, 4 誌であった。また、工学など 6 分野は 0~2 誌であった。全 220 誌の掲載率はリポジトリが 59.5%、補足資料が 89.5%であった。研究倫理や COI (利益相反) に関する記述がある雑誌や商業出版社の雑誌は掲載率や強度が高い傾向にあることから、データ共有ポリシーは研究不正と関連があることが示唆された。

Data sharing policies in scholarly journals across different disciplines: A comparative analysis

Ui Ikeuchi (*Graduate School of Library, Information and Media Studies, University of Tsukuba*)

Hiroshi Isumura (*Faculty of Library, Information and Media Science, University of Tsukuba*)

【Abstract】

The purpose of this study is to examine the journal policies for research data sharing across a wide range of disciplines. The authors selected 22 subject areas and reviewed “Guidelines for authors” in 10 journals in each of these areas. We focused on two types of data sharing policies: (1) repository policy, where researchers deposit their data to repositories and indicate the identifier in their article, and (2) supplemental materials policy. We classified the policies into four ranks based on their strengths.

In most of the subject areas in the Biological and Medical Sciences, journals tend to have a high ratio of and strong repository policies, and they recommend the same data repositories. However, only seven journals in Agricultural Sciences, six in Pharmacology & Toxicology, and two in Psychiatry/Psychology have a repository policy. Although Geosciences, Space Science, and Social Sciences are the leading areas in data sharing, just seven, six, and four journals, respectively, have a repository policy. Only 0–2 journals have a repository policy in six other disciplines, such as Engineering. Of the 220 selected journals, 59.5% have a repository policy, whereas 89.5% have a supplemental materials policy. Journals having a description of research ethics or a disclosure for conflict of interest (COI), or published by commercial publishers tend to have a high ratio of and stronger policies. These results suggest that data sharing policies in scholarly journals are associated with scientific misconduct.

【目次】

1. はじめに	3
1.1 データ共有の進展と分野間の差異	3
1.2 学術雑誌によるデータ共有の推進	3
1.3 研究課題と用語の定義	4
2. 先行研究	4
3. 調査・分析方法	5
3.1 分野別学術雑誌の選定	5
3.2 ポリシーの強度と公開先リポジトリ	6
3.3 雑誌の特徴	6
3.4 分析方法	7
4. 結果	7
4.1 データ共有ポリシーと雑誌の特徴の概要	7
4.2 分野別データ共有ポリシー	9
4.2.1 掲載率が高い分野（グループ A）	9
4.2.2 掲載率が中程度の分野（グループ B）	9
4.2.3 掲載率が低い分野（グループ C）	10
4.3 調査対象誌の特徴	10
4.3.1 倫理と COI（利益相反）の開示	11
4.3.2 出版者：学協会と商業出版社	12
4.3.3 OA ステータス	13
4.3.4 出版国	13
4.3.5 インパクト指標	14
4.3.6 刊行開始後年	15
4.4 ポリシーの強度に影響を与える特徴	15
5. 考察	16
5.1 分野別データ共有ポリシーの状況と展望	16
5.2 雑誌の特徴とデータ共有ポリシー	17
5.3 データ共有ポリシーに関する知見と活用	18
6. おわりに	18
謝辞	19
注・引用文献	19

1. はじめに

1.1 データ共有の進展と分野間の差異

科学研究の成果として、論文のみならずデータをも共有する動きが盛んである¹²⁾。その背景には、ICT（情報通信技術）の発達やサイバーインフラストラクチャの普及があり³⁴⁾、データジャーナルの刊行や学術雑誌へのデータペーパーの掲載も増加している⁵⁾。データ共有の主な目的は、後続研究の効率化や研究開発費の節減⁶⁾、研究の信頼性の向上⁷⁸⁾であり、科学コミュニティだけではなく社会全体にとっても意義がある⁹⁾¹⁰⁾。そのため各国の政府や研究助成機関（以下、「助成機関」とする）によるデータ共有の義務化が相次ぎ¹¹⁾、2007年には「公的資金による研究データへのアクセスに関する OECD 原則とガイドライン（以下、OECD 原則）」¹²⁾が、2013年には EC（欧州委員会）のガイドライン¹³⁾が公開されるなど、データ共有の推進は国際的な政策課題となりつつある。

さらに、異分野データの統合による新たな成果の創出¹⁴⁾や共同研究の推進¹⁵⁾が期待されており、分野を超えたデータ共有の実現を目指して RDA（Research Data Alliance）¹⁶⁾や DataCite¹⁷⁾といった国際組織が基盤整備を進めている。また、研究データの公開に関する G8 の科学大臣共同声明では“社会科学、人文科学、自然科学、生命科学、環境科学にわたる学際的で分野横断的なアプローチを取ることの重要性”が強調されている¹⁸⁾。しかし、現状ではデータを共有するためのリポジトリ数は分野によって異なり（表 1）、Christine L. Borgman は“天文学やゲノムでしか有効に活用できる状態で公開されていない”¹⁹⁾と指摘している。そこで本研究は、機密やプライバシーに触れない限りあらゆる分野のデータが共有されることが望ましいとする立場から、共有が遅れている分野を明らかにすることによって適切な支援を可能にし、異分野融合による科学の進展に寄与することを目指す。

表 1 Databib²⁰⁾に登録された分野別データリポジトリ数

分野	リポジトリ数
学際分野	179 (18.0%)
生物科学	174 (17.5%)
地球科学	117 (11.8%)
数学・物理学	106 (10.7%)
社会科学	92 (9.3%)
環境科学	87 (8.8%)
健康・医学	60 (6.0%)
農業	18 (1.8%)
生態系科学	15 (1.5%)
言語学・文学	11 (1.1%)
その他(工学, ビジネスほか)	48 (4.8%)
未分類	85 (8.6%)
合計	992 (100%)

2014年12月24日現在

1.2 学術雑誌によるデータ共有の推進

分野によるデータ共有の推進状況を比較するため、本研究は学術雑誌（以下、雑誌）のデータ共有ポリシーを調査する。雑誌のポリシーを対象とする理由は、(1)雑誌によるデータ共有が有用であると認識されていること、(2)雑誌のポリシーによる要求がデータ共有の動機となっていること、(3)要求が強いほ

どデータの公開率が高いことである。まず(1)について、雑誌によるデータ共有、すなわちリポジトリで公開したデータの識別子を論文に記載することや、補足資料としてデータを掲載することの重要性が指摘されており²¹⁾²²⁾、出版者にポリシーの策定を求める提言が行われている²³⁾²⁴⁾。リンク切れ²⁵⁾や引用方法が未整備である²⁶⁾といった問題も指摘されているが、生命科学分野の研究者を対象とした調査では、研究の推進に最も役立つとされたデータ共有ツールは「雑誌の補足データ」(58%)であった²⁷⁾。(2)の共有動機について、研究者がデータを共有する際には、技術やコスト、機密といった問題が生じる²⁸⁾。しかし、雑誌のポリシーによるデータ共有の要求が、こうした障壁を超える動機となっている²⁹⁾。生物多様性データの共有に関する調査によれば、外的な共有動機の1位は「雑誌の要求」(59%)であった³⁰⁾。(3)について、Timothy H. Vinesらは要求が強い雑誌ほどデータの公開率が高いことを示し、ポリシーがデータ共有を強力に推進していると論じている³¹⁾。

1.3 研究課題と用語の定義

本研究の課題は次の2点とする。第一に、雑誌のデータ共有ポリシーを分野横断で調査することによって、分野別のデータ共有に対する姿勢や推進状況を比較すること、第二に、データ共有ポリシーに影響を与える雑誌の特徴を分析することによって、今後の方向性を展望することである。

分野によってデータ共有の推進状況に差があることは繰り返し指摘されているが、客観的に差を示した研究はみられない。そこで世界的にデータ共有が推進されている現時点での分野別の状況を明らかにすること、また、後述する先行研究では未調査の分野を対象とすること、データ共有ポリシーに関連がある要因を明らかにすることが本研究の学術的貢献であると考えられる。社会的な意義としては、政策決定者や学術出版者の議論に資することを挙げる。分野間の差異が明らかになれば、効率的な学術情報政策の立案が可能となるであろう。多岐にわたるポリシーを調査する過程で得られた知見は、新たなポリシー策定のための基礎資料としても有益であると考えられる。

本研究における「データ」はOECD原則の定義³²⁾に準じて、科学研究の一次ソースとして利用される事実の記録(数値、動画など)で、コンピュータで読み取り可能なデジタル形式のものとする。雑誌による「データ共有」とは、データをリポジトリに公開してDOIやアクセッション番号などの識別子を論文に示す方法と、論文の補足資料として電子版の雑誌にデータや識別子を掲載する方法³³⁾の2種類とする。「データ共有ポリシー」とは、雑誌の投稿規定や執筆要綱など(以下、投稿規定)に掲載されたデータ共有に関する記述とする。

2. 先行研究

雑誌のデータ共有ポリシー(以下、ポリシー)に関する先行研究を表2に示す。2007年にHeather A. Piiowarらは、マイクロアレイによる遺伝子発現解析に関する論文の掲載誌70誌を調査して、42誌がポリシーを掲載していることを明らかにした³⁴⁾。2009年にAlawi A. Alsheikh-Aliらは、JCR(Journal Citation Reports)を用いてIF(Impact Factor)上位50誌を対象とした調査を行った。調査対象誌は生命科学や学際分野であり、44誌が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ポリシーを掲載していることを明らかにした³⁵⁾。2011年と2012年にVictoria Stoddenらは、確率統計、数理・計算生物学など計算科学分野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調査した。その結果、データのポリシーを掲載する雑誌は一年間で56誌から64誌に増加していた³⁶⁾。2013年にSven Vlaeminckが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誌にデータの可用性(availability)ポリシーが示されていた³⁷⁾。2014年にWolfgang Zenk-Möltgenらが発表した論文によれば、SSCI(Social Science Citation Index)に収録された社会学分野の雑誌140誌のうち、101誌にポリシーが掲載されていた³⁸⁾。

以上の先行研究は対象分野が限定されており調査時期も異なるため、本研究は分野横断の調査を同時に実施した上で分野間の比較を行う。

表 2 学術雑誌のデータ共有ポリシーに関する先行調査の概要

著者	調査年	調査対象	雑誌数	掲載誌数
Piwowar ら	2007	生命科学	70	42 (60.0%)
Alsheikh-Ali ら	2009	IF 上位 50 誌 (52.6~8.7) 生命科学等	50	44 (88.0%)
Stodden ら	2011	計算科学	170	56 (32.9%)
	2012		170	64 (37.6%)
Vlaeminck	2013 [†]	経済学	141	29 (20.6%)
Zenk-Möltgen ら	2014 [†]	社会学	140	101(72.1%)

[†]論文の発表年

Piwowar らと Zenk-Möltgen らはポリシーを、それぞれ「強い」「弱い」「なし」と「あり」「奨励」「なし」の 3 段階の強度に、Stodden らは後述する 5 段階の強度に分類した。また、ポリシーと関連する雑誌の特徴を把握するために、出版者、OA (Open Access) ステータス、IF を調査している。出版者について、Piwowar らの調査では、学協会が 82%、商業出版社は 50%がポリシーを掲載しており、Stodden らの予測モデルでも学協会の方がポリシーを掲載する確率が高かった。また、Piwowar らの調査では OA の 4 誌は全てポリシーを掲載しており、重回帰分析の結果 OA は IF とともにマイクロアレイデータの共有ポリシーの強度に影響を与えていた。一方、Stodden らの調査では、OA 誌はポリシーの掲載率が高かったものの統計的な有意差はみられなかった。Piwowar らはポリシーの強度が高いほど IF の中央値が高いことを示し、Stodden らの予測モデルでは IF が高い雑誌ほどポリシーを掲載する確率が高かった。Zenk-Möltgen らの調査でも、5 年 IF の値が高いほどポリシーの掲載率が有意に高かった。また、Stodden らの結果、および Piwowar らと Alsheikh-Ali らの結果の比較から、調査年が新しいほどポリシーの掲載率が高いといえる。本研究はこれらに加えて、出版国などポリシーに影響を与えると考えられる特徴についても調査する。

3. 調査・分析方法

先行研究を参考に、2013 年 10 月 1 日から 30 日に予備調査³⁹⁾を実施して調査・分析方法を決定した。投稿規定や特徴は、雑誌のウェブサイトと Ulrichweb.com で確認した。ウェブサイトは 2014 年 4 月 8 日から 5 月 8 日に、Ulrichweb.com は 2014 年 6 月 5 日に調査した。なお、全ての調査対象誌は電子版が電子版と印刷版の両方が存在しており、投稿規定を雑誌のウェブサイトで確認することができた。

3.1 分野別学術雑誌の選定

分野の分類には、研究動向・統計データベースである ESI (Essential Science Indicators) の 22 分野を用いた。予備調査は 18 分野各 10 誌を対象としたが、医学の下位分野でポリシーの掲載率や強度に差がみられた。そこで、医学が細分されていること、かつデータ共有に関する文献⁴⁰⁾で取り上げられている分野を含んでいることから ESI に決定した。また、先行研究と予備調査から IF が低い雑誌の多くはポリシーが掲載されていない可能性が高く⁴¹⁾、分野間比較が困難であると予想されることから、調査対象は各

分野のIFが高い雑誌とした。調査に用いた2013年9月版のESI雑誌リスト⁴²⁾には11,872誌が収録されており、各雑誌に自然科学と人文・社会科学の22分野が一意に付与されている。まず、ISSNがなく2012年版のJCRにも収録されていなかった27タイトルを除いて11,845誌を抽出した。続いて、IFをJCR(2012)から取得して、分野ごとに順位付けを行った。

さらに、調査対象を原著論文を中心に掲載している雑誌（以下、原著論文誌）に限定した。予備調査の過程でレビュー誌、速報誌、短報誌、コメンタリー誌は、その性質上データをほとんど公開していないことが明らかになったため、雑誌の概要や掲載論文を参考に原著論文誌以外を除外した。

以上の手続きによって、22分野のIF上位10誌、合計220誌を抽出して調査対象とした。

3.2 ポリシーの強度と公開先リポジトリ

ポリシーの要求の強度は、(1)データ共有を要求している場合や査読・出版の条件としている場合は「必須」、(2)データ共有を奨励している場合は「推奨」、(3)データを受け付ける／共有について何らかの記述がある場合は「受諾」、(4)全く言及されていない場合は「なし」の4段階に分類した（表3）。予備調査では、Stoddenらと同様に「必須」を「査読・出版条件として要求」と「要求」に分類して5段階の分類を試みたが、“要求する”とだけ書かれていたとしてもデータを公開しなければ査読・出版されない可能性があるため、4段階の分類とした。

表3 データ共有ポリシーの要求の強度の分類

要求の強度	表現例
必須	require, must, should, condition of publication
推奨	encourage, recommend
受諾	accept, allow, might, possible to include
なし	(記載なし)

また、永続性のあるリポジトリはデータ共有の基盤とされている⁴³⁾ことから、公開先リポジトリの例示の有無とリポジトリ名も調査した。

3.3 雑誌の特徴

先行研究を参考に、以下の6項目を調査した。(1), (3)は雑誌のウェブサイトで調査し、(2), (4), (6)はUlrichweb.comを用いた。(5)は後述する。

(1) 倫理とCOI（利益相反）の開示に関する記述

データ共有の目的の一つに研究不正への対策があることから「研究倫理やCOIの開示に関する記述の有無はポリシーと関連がある」という仮説を立てて調査した。

(2) 出版者

先行研究と同様に、出版者を学協会と商業出版社に分類した。商業出版社から学協会の雑誌が刊行されている場合は、出版者を学協会、所在地を出版国とした。

(3) OAステータス

OAステータスは急速に変化している⁴⁴⁾ため、既存のデータベースは使用せず、雑誌のウェブサイトでも調査して「OA」、「制約つきOA」、「非OA」に分類した。「制約つきOA」には、著者が追加のAPC

(論文出版加工料)を支払うことによって特定の論文を OA とするハイブリッド型の雑誌や、エンバゴを要する雑誌を含めた。

(4) 出版国

各国の政府や助成機関がデータ共有を義務化していることから、出版国を調査した。データ共有に関連する各国の数値情報として、データリポジトリのディレクトリである re3data.org⁴⁵⁾に収録されている認証リポジトリ⁴⁶⁾数を調査した。また、政府のオープンデータと研究データの共有は、市民へのデータ公開によって政策や研究の透明性を向上させるという同質の目的があり⁴⁷⁾、いずれも公共財として広く共有されるべきであると論じられてきた⁴⁸⁾⁴⁹⁾ことから、ODI (Open Data Index) 2014 年版⁵⁰⁾による政府データの公開率を調査した。

(5) インパクト指標

雑誌の収録論文の被引用回数に基づく「インパクト指標」の違いによる差を明らかにするため、調査時点で最新版の IF と Google Scholar Metrics h-index (以下、GSM)⁵¹⁾の値を調査した。IF は 2012 年度版の JCR を、GSM は 2013 年 7 月版を用いた。

(6) 刊行開始年

先行研究では調査年が新しいほどポリシーの掲載率が高かったことから、「刊行開始から経過した年数(以下、刊行開始後年)の長短はポリシーと関連がある」という仮説を立てて調査した。

3.4 分析方法

雑誌の特徴とポリシーの掲載率および強度の関連を明らかにするため、以下の分析を行った。数値同士の相関を調べる際は Pearson の相関係数を、ポリシーの強度との相関は、強度を 4 (必須), 3 (推奨), 2 (受諾), 1 (なし) の 4 段階の順序尺度として、Spearman の順位相関係数を用いた。

雑誌の特徴のうち、倫理と COI, 出版者, OA ステータス, 出版国はポリシーの強度とのクロス集計表を作成し、カイ二乗検定を行った。期待度数が 5 未満になるセルが表全体の 20% を超える場合は、隣接するセルの値を合算した。また、調整済み残差が 2 を超える場合は全体の傾向と差があるとみなした。インパクト指標と刊行開始後年はポリシーの強度との相関を調べた。インパクト指標は強度ごとの IF と GSM の中央値も比較した。

以上の結果をふまえて、どのような特徴をもつ雑誌のポリシーの強度が高いのかを明らかにするために順序ロジスティック回帰分析を行った。目的変数を強度として、特徴から説明変数を選択した。ソフトウェアは SPSS Version22 を用いた。

4. 結果

4.1 データ共有ポリシーと雑誌の特徴の概要

全 220 誌のうち、204 誌 (92.7%) がリポジトリまたは補足資料によるポリシーを掲載していた。掲載率は、本研究よりも IF が高い雑誌を対象とした Alsheikh-Ali らを含む表 2 の先行研究を上回った。

表 4 に、分野別の集計結果をリポジトリによるポリシーの掲載誌が多い順 (同数の場合は表 3 の強度が高い順) に示す。リポジトリによるポリシーは 131 誌 (59.5%), 補足資料は 197 誌 (89.5%) が掲載しており、両方を掲載している雑誌は 124 誌 (56.4%) であった。リポジトリは強度が高い「必須」と「推奨」が全体の 59.1% を占めるのに対して、補足資料は強度が低い「受諾」が 62.3% であった。公開先リポジトリを指示していた 124 誌のうち、121 誌は具体的にリポジトリ名を例示していたが、*Retrovirology* 等 3 誌は既存のディレクトリ (DataCite, re3data.org または Databib) から選ぶよう指示していた⁵²⁾。また、雑誌の特徴として倫理, COI の記述がある雑誌数, 学協会, OA と制約つき OA の雑誌数, IF, GSM, 刊行開始後年の平均値を示した。詳細は 3.3 で述べる。

表 4 分野別データ共有ポリシーと雑誌の特徴

		リポジトリ				補足資料				公開先 リポジトリ	雑誌の特徴							
		必須	推奨	受諾	なし	必須	推奨	受諾	なし		倫理	COI	学協会	OA [†]	IF	GSM	year	
A	1	分子生物学・遺伝学	10	0	0	0	3	0	7	0	10	7	8	2	8	18.0	120.2	24.1
	2	生物学・生化学	9	1	0	0	5	1	4	0	10	9	10	2	5	16.1	88.2	19.4
	3	臨床医学	9	1	0	0	2	0	8	0	10	4	9	4	6	27.5	164.4	67.2
	4	微生物学	8	0	1	1	1	2	7	0	9	8	8	3	7	6.5	63.8	15.4
	5	材料科学	8	0	0	2	3	1	6	0	8	8	5	2	7	14.0	105.2	16.7
	6	植物学・動物学	8	0	0	2	2	0	7	1	8	4	4	8	7	6.4	56.2	41.5
	7	化学	7	1	0	2	3	4	3	0	8	8	8	8	4	10.7	87.0	38.2
	8	神経科学・行動科学	7	1	0	2	1	1	7	1	6	6	8	4	8	11.0	98.1	46.5
	9	環境科学・生態学	5	3	0	2	2	2	6	0	8	8	8	5	6	9.9	61.1	28.9
	10	免疫学	5	3	0	2	1	1	8	0	8	9	10	4	8	10.3	83.2	47.8
B	11	学際分野	7	0	0	3	4	1	4	1	7	6	9	6	6	10.3	111.6	68.3
	12	農学	3	4	0	3	1	1	7	1	6	6	6	5	10	3.7	45.1	39.1
	13	地球科学	2	5	0	3	1	1	8	0	7	6	4	7	8	6.0	53.9	35.4
	14	薬理学・毒物学	4	2	0	4	0	2	8	0	6	8	9	5	6	5.9	51.6	35.1
	15	宇宙科学	1	5	0	4	0	5	4	1	6	3	1	8	4	6.3	79.6	82.3
	16	社会科学	2	2	0	6	0	2	5	3	2	4	6	7	7	4.4	36.2	47.5
C	17	物理学	2	0	0	8	2	2	6	0	2	7	7	6	7	9.7	89.2	34.5
	18	コンピュータサイエンス	2	0	0	8	0	1	8	1	2	3	4	6	8	4.8	50.7	22.5
	19	精神医学・心理学	1	1	0	8	0	1	7	2	1	6	8	9	6	8.2	62.3	72.1
	20	数学	0	1	0	9	0	1	2	7	0	3	4	8	5	3.7	37.7	63.5
	21	工学	0	0	0	10	0	0	9	1	0	3	3	5	4	5.3	62.4	39.0
	22	経済学・ビジネス	0	0	0	10	0	0	6	4	0	3	4	7	3	5.1	57.3	51.1
合計		100	30	1	89	31	29	137	23	124	129	143	121	140	(9.3)	(75.7)	(42.6)	

N=220, GSM のみ N=214 †OA は制約つき OA 誌を含む IF, GSM, year (刊行開始後年) は平均値

多くの雑誌で例示されていたリポジトリの上位 10 件を表 5 に示す。なお、こうしたリポジトリ名を例示した上で、*Molecular Ecology Resources* (環境科学・生態学) は機関リポジトリか助成機関のリポジトリで、*Genome Research* (分子生物学・遺伝学) は著者のウェブサイトでデータを公開することも認めていた (各 1 誌のみ)。

表 5 データの公開先として例示されたりポジトリ (上位 10 件)

リポジトリ名	雑誌数	開始年	主なデータ
GenBank [†]	79	1983	塩基配列
GEO (Gen Expression Omnibus)	64	2002	遺伝子発現
ArrayExpress	59	1994	遺伝子発現
DDBJ [†]	59	1986	塩基配列
EMBL [†]	56	1980	塩基配列
PDB (Protein Data Bank)	50	1971	蛋白質
UniProt/Swiss-Prot/PIR	46	2002	蛋白質
CSD (Cambridge Structural Database)	40	1987	結晶構造
PANGAEA	22	1996	地球科学・環境科学
Dryad	19	2008	自然科学・医学

[†]GenBank, DDBJ, EMBL は INSDC (統一フォーマットの国際塩基配列データベース) で相互運用されているが、いずれか 1 つ、2 つ、あるいは全てを例示する雑誌がみられたため、それぞれ集計した。

4.2 分野別データ共有ポリシー

本節では、リポジトリによるポリシーを掲載している雑誌の合計が 10~8 誌の 10 分野を A, 7~4 誌の 6 分野を B, 3~0 誌の 6 分野を C として、それぞれの調査結果を述べる。

4.2.1 掲載率が高い分野 (グループ A)

グループ A には、分子生物学や臨床医学など広義の生物・医学分野、および材料科学、化学、環境科学・生態学が含まれている。表 5 に示した GenBank や GEO など生物データのリポジトリは、グループ A の 10 分野各 10 誌のいずれかで必ず例示されていた。さらに、材料科学と化学分野は CSD などの結晶構造のリポジトリを、環境科学・生態学は PANGAEA や Dryad を例示している。表 5 に挙げたりポジトリは、多くの機関による認証・評価基準を満たしている⁵³⁾⁵⁴⁾。つまり、この 10 分野はデータの公開先として科学コミュニティに認知された「定評のあるリポジトリ」が存在する。また、10 分野各 10 誌の全てにリポジトリか補足資料どちらかのポリシーが掲載されていた。

4.2.2 掲載率が中程度の分野 (グループ B)

地球科学、宇宙科学、社会科学はデータ共有が盛んであるとされているが、ポリシーの掲載率や強度は中程度であった。一方で、研究対象とした一次データを特定するためのリンクの記述などを求めている雑誌が多い。たとえば宇宙科学の *The Astrophysical Journal* など 3 誌は、NASA Archive にデータがあればリンクするよう指示しており、社会科学の *Health Reports* は社会統計などの元データへの記述を含まなければならないとしていた。つまり、この 3 分野では共有されたデータの二次分析が行われる⁵⁵⁾ため、リポジトリによるポリシーの掲載率や強度が中程度であると考えられる。

学際分野、農学、薬理学・毒物学のリポジトリによるポリシーは、全て表 5 の生物データリポジトリを例示していたが、他の生物・医学分野と比較して掲載率や強度が低かった。

グループBでリポジトリと補足資料いずれのポリシーもなかったのは、農学、宇宙科学、社会科学の各1誌、合計3誌であった。

4.2.3 掲載率が低い分野（グループC）

物理学と数学は、表1に示したDatabibの収録リポジトリ数が106件であり比較的多い。物理学は、リポジトリによるポリシーは「必須」2誌のみであったが、「なし」の雑誌であってもデータに関する記述がみられた。たとえば、*European Physical Journal C*にはデータの改竄について言及していた。一方、数学は「推奨」1誌のみであり、当該誌のDatabaseはデータベースと情報源を少なくとも2年間アクセス可能にするよう推奨しているものの、公開先を指示していない。数学は補足資料も7誌が「なし」であり、22分野で最も掲載率が低い。物理学の対象誌は実験や観察によるデータ作成や収集を伴う研究を含んでいたが、数学の対象誌はデータを用いない理論研究が多かったことが両分野の差につながったと考えられる。

コンピュータサイエンス（以下、CS）、工学、経済学・ビジネスはDatabibの収録リポジトリ数が少ない（表1）。この3分野でリポジトリによるポリシーを掲載していたのはCSの2誌のみであったが、研究に用いたデータに関する記述は各分野でみられた。たとえば、IEEEの雑誌はCSに4誌、工学に5誌含まれているが、提出するデータの形式についてFAQで言及している。また、経済学・ビジネスの*Academy of Management Journal*は、論文に使われたデータが過去にも使われたかどうかを明らかにするよう求めている。つまり、この3分野はデータを用いた研究が行われているにもかかわらず、ポリシーの掲載率が低かった。

精神医学・心理学はデータ共有の進展を目指す分野であるが⁵⁰、リポジトリによるポリシーは「必須」1誌、「推奨」1誌のみであった。*British Journal of Psychiatry*はデータの公開を“強く推奨”しているものの公開先を指定していなかった。

グループCでリポジトリと補足資料いずれのポリシーも掲載していなかったのは、数学が6誌、経済学・ビジネスが4誌、精神医学・心理学が2誌、工学が1誌で合計13誌であった。

4.3 調査対象誌の特徴

本節では、雑誌の特徴とポリシーの掲載率および強度についての結果を示す。リポジトリによるポリシーは「受諾」が1誌であったため、カイ二乗検定の際はリポジトリのみ「受諾」を「推奨」と合算した。また、表4に示したように雑誌の特徴は分野による差がみられた。有意水準5%で二項検定を行ったところ、倫理は2分野、COIは6分野、学協会は5分野が母集団から外れていたため、OA（1分野）以外は分野によって特徴の分布が大きく異なるといえる。そこで倫理、COI、学協会は全220誌の結果を示した後に、分野による交絡を制御するため生物・医学と自然科学の「分野グループ（表6）」に分けた結果を示す。

表6 分野グループ

生物・医学分野	自然科学分野	その他
分子生物学・遺伝学、生物学・生化学、臨床医学、微生物学、植物学・動物学、神経科学・行動科学、免疫学、農学、薬理学・毒物学、精神医学・心理学	材料科学、化学、環境科学・生態学、地球科学、宇宙科学、物理学、コンピュータサイエンス、数学、工学	学際分野、社会科学、経済学・ビジネス
10 分野	9 分野	3 分野

4.3.1 倫理と COI（利益相反）の開示

倫理に関する記述を含む雑誌は 129 誌であり、その主な内容はデータの改竄や捏造、オーサーシップといった出版倫理に関する規定であった。COI の開示に関する記述を含む雑誌は 143 誌であり、資金提供や共同研究者など研究に影響を与えると考えられる利害関係の事実を開示するよう求めていた。国際的な指針に準拠している例も多く、CONSORT（臨床試験報告に関する統合基準）声明⁵⁷⁾は 38 誌、COPE（出版倫理委員会）ガイドライン⁵⁸⁾は 27 誌、ICMJE（医学雑誌編集者国際委員会）による勧告⁵⁹⁾は 18 誌で採用されていた。

ポリシーの掲載率は、倫理の記述有りの雑誌は 129 誌中 126 誌で 97.7%、無しは 91 誌中 78 誌で 85.7%、COI 有りの雑誌は 143 誌中 139 誌で 97.2%、無しは 77 誌中 65 誌で 84.4%であった。リポジトリ・補足資料それぞれのポリシー掲載率も、倫理・COI 有りの雑誌の方が無しの雑誌よりも高かった（表 7）。強度をみると、倫理・COI とともに有りの雑誌はリポジトリ・補足資料によるデータ共有を「必須」とする雑誌の割合が多く、「なし」の割合が少ない。倫理・COI の有無とリポジトリ・補足資料の強度の組み合わせのカイ二乗検定の結果は、全て有意差がみられた（4 組とも $P < 0.01$ ）。

表 7 倫理・COIに関する記述の有無とデータ共有ポリシー

		雑誌数	リポジトリ				補足資料			
			必須	推奨	受諾	なし	必須	推奨	受諾	なし
倫理	有	129	68	22	1	38	25	19	78	7
	無	91	32	8	0	51	6	10	59	16
COI	有	143	78	20	1	44	30	16	89	8
	無	77	22	10	0	45	1	13	48	15
合計		220	100	30	1	89	31	29	137	23

表 8 に分野グループ別の結果を示す。倫理有りは生物・医学が 67.0%、自然科学が 54.4%であり、COI 有りは生物・医学が 80.0%、自然科学は 48.9%で、いずれも生物・医学の掲載率が高かった。

倫理について、リポジトリ・補足資料によるポリシーの掲載率は、生物・医学、自然科学ともに記述有りの方が無しよりも高かった。特に自然科学で無しの雑誌はリポジトリによるポリシーの掲載率が低く、12 誌 (29.3%) にとどまった。強度も自然科学はリポジトリ・補足資料とともに倫理の有無で「必須」と「なし」の差が大きく有意差がみられた（いずれも $P < 0.01$ ）。生物・医学では、リポジトリの「推奨」と「受諾」の合計が有りは 13 誌、無しは 1 誌であり差が大きかった。

COI について、リポジトリ・補足資料によるポリシーの掲載率は、生物・医学、自然科学ともに記述有りの方が無しよりも高く、特に自然科学で無しの雑誌はリポジトリによるポリシーが 18 誌 (39.1%) にしか掲載されていなかった。強度については自然科学の補足資料に有意差がみられ ($P < 0.05$)、特に「必須」の差が大きかった。

表 8 倫理・COIに関する記述の有無とデータ共有ポリシー（分野グループ別）

			雑誌数	リポジトリ				補足資料			
				必須	推奨	受諾	なし	必須	推奨	受諾	なし
倫理	生物・医学	有	67	42	12	1	12	13	8	43	3
		無	33	22	1	0	10	3	1	27	2
	自然科学	有	49	22	8	0	19	10	10	28	1
		無	41	5	7	0	29	1	7	24	9
COI	生物・医学	有	80	54	11	1	14	16	6	56	2
		無	20	10	2	0	8	0	3	14	3
	自然科学	有	44	17	7	0	20	10	8	24	2
		無	46	10	8	0	28	1	9	28	8
合計			190	91	28	1	70	27	26	122	15

4.3.2 出版者：学協会と商業出版社

ポリシーの掲載率は、学協会が121誌中109誌で90.1%、商業出版社が99誌中95誌で96.0%であった。リポジトリによるポリシーの掲載率は、学協会が52.9%、商業出版社が67.7%、補足資料は学協会が86.0%、商業出版社が93.9%であり、全て商業出版社の掲載率の方が高かった。強度をみると、リポジトリによるデータ共有を「必須」とする雑誌の割合は学協会38.8%、商業出版社53.5%、補足資料「必須」は学協会9.1%、商業出版社20.2%であった。リポジトリ「なし」は学協会47.1%、商業出版社32.3%、補足資料「なし」は学協会14.0%、商業出版社6.1%と商業出版社の比率が低く、補足資料については有意差もみられた ($P<0.05$)。以上の結果から、商業出版社の雑誌は学協会の雑誌に比べてポリシーの掲載率と強度が高いという傾向がみられた。

表 9 学協会・商業出版社別データ共有ポリシー

出版者	雑誌数	リポジトリ				補足資料			
		必須	推奨	受諾	なし	必須	推奨	受諾	なし
学協会	121	47	17	0	57	11	20	73	17
商業出版社	99	53	13	1	32	20	9	64	6
合計	220	100	30	1	89	31	29	137	23

表10に分野グループ別の結果を示す。学協会の雑誌は生物・医学は46.0%、自然科学は61.1%であった（全体は53.2%）。リポジトリ・補足資料によるポリシーの掲載率は、分野グループ別でも全て商業出版社の方が高かった。強度は自然科学の補足資料が全体の傾向と異なっており、「推奨」が学協会は27.3%と多く、商業出版社は5.7%で少なかった ($P<0.05$)。「必須」とあわせても学協会34.5%、商業出版社25.7%であり、学協会の強度が高い。従って、自然科学は学協会の方が補足資料のポリシーの強度が高かった。

表 10 学協会・商業出版社別データ共有ポリシー（分野グループ別）

	出版者	雑誌数	リポジトリ				補足資料			
			必須	推奨	受諾	なし	必須	推奨	受諾	なし
生物・医学	学協会	46	27	6	0	13	5	3	34	4
	商業出版社	54	37	7	1	9	11	6	36	1
自然科学	学協会	55	15	10	0	30	4	15	29	7
	商業出版社	35	12	5	0	18	7	2	23	3
合計		190	91	28	1	70	27	26	122	15

4.3.3 OA ステータス

Piwozar らの結果と同様に、全ての OA 誌はリポジトリか補足資料のいずれか、あるいは両方のポリシーが掲載されていた。OA と制約つき OA を合算して強度と OA ステータスのカイ二乗検定を行ったが、リポジトリによるポリシーの強度は OA ステータスと有意な関連がみられなかった。補足資料の強度は非 OA が高く、OA と制約つき OA は低いという傾向がみられた ($P<0.05$)。

表 11 OA ステータス別データ共有ポリシー

	雑誌数	リポジトリ				補足資料			
		必須	推奨	受諾	なし	必須	推奨	受諾	なし
OA	15	9	1	0	5	1	5	5	4
制約つき OA	125	54	23	0	48	11	14	91	9
非 OA	80	37	6	1	36	19	10	41	10
合計	220	100	30	1	89	31	29	137	23

4.3.4 出版国

調査対象誌の出版国は 13 か国であった⁶⁰。上位 4 か国、合計 201 誌（全体の 91.4%）の集計結果を表 12 に示す。ポリシーの掲載率は、ドイツとオランダが 100%、英国が 97.2%、米国が 87.5%であった。強度について、リポジトリによるポリシーが「必須」の割合は英国 54.9%、ドイツ 53.8%、米国 47.1%、オランダ 15.4%の順であり、補足資料が「必須」の割合も英国 23.9%、ドイツ 15.4%、米国 11.5%、オランダ 0%と同じ順位であった。また、4 か国とりポジトリ・補足資料の強度には有意差がみられた（どちらも $P<0.01$ ）。

13 か国の ODI データ公開率とデータ共有「必須」の割合の関連を調べたところ、リポジトリ「必須」との相関はみられなかったが、補足資料「必須」とは 0.60 ($P<0.05$) と比較的高い相関がみられた。同じく re3data.org の認証リポジトリ数とりポジトリ「必須」の相関は 0.71 ($P<0.01$)、補足資料「必須」との相関は 0.64 ($P<0.05$) であった。

表 12 出版国別データ共有ポリシー（上位4か国, N=201）

国名	雑誌数	リポジトリ				補足資料				ODI [†]	re3 [‡]
		必須	推奨	受諾	なし	必須	推奨	受諾	なし	公開率	認証
米国	104	49	6	0	49	12	12	63	17	70%	88
英国	71	39	13	1	18	17	17	34	3	97%	25
ドイツ	13	7	2	0	4	2	0	11	0	69%	60
オランダ	13	2	4	0	7	0	0	12	1	64%	15

[†]2014 年度版 Open Data Index によるデータ公開率

[‡]2015 年 6 月 25 日現在 re3data.org に収録されている認証リポジトリ数

4.3.5 インパクト指標

Piwovar らはポリシーを「強い」、「弱い」、「なし」に分類した上で、強度が高いほど IF の中央値が高いことを示した。本研究の「推奨」と「受諾」は「弱い」に相当すると判断して、それぞれの IF および GSM の中央値を求めた結果、リポジトリ・補足資料によるポリシーの強度が高い雑誌ほど IF・GSM の中央値が高かった（表 13）。

表 13 データ共有ポリシーの強度とインパクト指標の中央値

	IF			GSM	
	Piwovar ら	リポジトリ	補足資料	リポジトリ	補足資料
強い（必須）	6.0	10.2	14.1	90.0	108.0
弱い（推奨, 受諾）	4.5	5.9	6.2	60.0	65.5
なし	3.6	4.8	4.3	46.0	39.0
全体	4.5		6.4		63.0
雑誌数	70		220		214

生物・医学は Piwovar らの対象分野に近いと考えられるが、IF はリポジトリの「弱い」が 5.9, 「なし」が 6.0 であり結果が異なった。その他は全て、強度が高いほどインパクト指標の中央値が高かった（表 14）。

表 14 データ共有ポリシーの強度とインパクト指標の中央値（分野グループ別）

	生物・医学				自然科学			
	IF		GSM		IF		GSM	
	リポジトリ	補足資料	リポジトリ	補足資料	リポジトリ	補足資料	リポジトリ	補足資料
強い	11.6	14.1	91.5	101.5	9.0	14.5	85.5	108.0
弱い	5.9	7.6	55.0	71.0	6.3	6.0	69.5	65.0
なし	6.0	6.2	44.0	56.5	4.8	3.9	47.0	36.5
全体		8.1		77.0		6.1		58.0
雑誌数		100		97		90		88

分野の影響を排してインパクト指標と強度の関連をみるため、分野ごとに IF と GSM が高い順に 10

から1の順位をつけて相関分析を行った。しかし、IFとリポジトリは0.16 ($P<0.05$)、IFと補足資料は0.24 ($P<0.01$)、GSMとリポジトリは0.18 ($P<0.01$)であり、いずれも低い相関しか認められず、GSMと補足資料は有意差がみられなかった。

4.3.6 刊行開始後年

刊行開始後年とポリシーの掲載の有無の相関分析の結果は、リポジトリが-0.17 ($P<0.01$)、補足資料が-0.15 ($P<0.05$)であり、ほとんど相関がみられなかった。刊行開始後年と強度もリポジトリが-0.20 ($P<0.01$)、補足資料が-0.24 ($P<0.01$)であり低い負の相関しか認められなかった。従って、最近刊行された雑誌の方がやや掲載率や強度が高いものの、強い関連はみられなかった。

4.4 ポリシーの強度に影響を与える特徴

ポリシーの強度に影響を与える特徴を明らかにするため、順序ロジスティック回帰分析を行った。(1)倫理の有無、(2)COIの有無、(3)出版者(商業出版社/学協会)、(4)OAステータス(OAと制約つきOA/非OA)、(5)IF、(6)GSM、(7)ODIデータ公開率、(8)認証リポジトリ数、(9)刊行開始後年に、分野による影響を制御するための(10)分野グループ(生物・医学)を追加して合計10変数とした。

まず、多重共線性が生じないように変数同士の相関分析を行った。カテゴリ変数は倫理とCOIに相関がみられたため(0.43, $P<0.01$)、他の変数との相関が高いCOIを変数から除外した。数値変数はIFとGSMに相関がみられたため(0.75, $P<0.01$)、線形回帰分析を行い、VIF値が高いGSMを除外した。なお、完全分離の状態にある変数はなかった。

以上の手続きによって8項目を説明変数、リポジトリと補足資料によるポリシーの強度を目的変数として順序ロジスティック回帰分析を実施した。説明変数の選択基準はWaldの χ^2 値が2以上かつ $P<0.05$ とした。4.3で強度と関連がみられた変数のうち、リポジトリは認証リポジトリ数と刊行開始後年、補足資料は出版者と認証リポジトリ数が選択基準を満たさなかったため除外した。原因として、強度との相関が高くないことや他の変数との交互作用が考えられる。最終的に、リポジトリはIF、生物・医学、倫理の3変数、補足資料はIF、ODIデータ公開率、倫理の3変数を選択した。Waldの χ^2 値が高いほど影響が強いと考えられる(表15)。補足資料について、刊行開始後年はODIデータ公開率と同様にIF、倫理との組み合わせでWaldの χ^2 値が2以上かつ $P<0.05$ であった。しかし、強度との相関やNagelkerkeの寄与率はODIデータ公開率の方が高かった。また、刊行開始後年とODIデータ公開率およびIFと倫理の4変数を投入すると、補足資料「なし」が有意にならない。以上の理由から、刊行開始後年は補足資料の説明変数として選択しなかった。

表 15 データ共有ポリシーの強度に影響を与える特徴 (N=220)

	偏回帰係数	標準誤差	Wald χ^2	有意確率
リポジトリ				
1 (なし)	2.260	0.387	34.142	0.000
2 (受諾)	2.287	0.387	34.823	0.000
3 (推奨)	3.089	0.415	55.304	0.000
IF	0.244	0.046	28.390	0.000
生物・医学	1.004	0.303	10.951	0.001
倫理 (有り)	0.740	0.303	5.958	0.015
補足資料				
1 (なし)	1.764	0.745	5.614	0.018
2 (受諾)	5.512	0.852	41.828	0.000
3 (推奨)	6.509	0.884	54.276	0.000
IF	0.086	0.018	22.332	0.000
ODI データ公開率	3.750	0.962	15.190	0.000
倫理 (有り)	1.112	0.305	13.279	0.000

リポジトリによるポリシーの強度が1 (なし) になる確率を Pr(1)とすると、回帰式は以下の通りになる。x1 は IF の値、x2 は生物・医学ならば1、その他は0、x3 は倫理有りは1、無しは0とする。

$$\text{logit}\{\text{Pr}(1)\} = 2.260 - 0.244x_1 - 1.004x_2 - 0.740x_3$$

$$\text{Pr}(1) = 1/\{1 + \text{Exp}(-2.260 + 0.244x_1 + 1.004x_2 + 0.740x_3)\}$$

同様に、表 15 から次の回帰式が得られる。

$$\text{logit}\{\text{Pr}(1 \text{ or } 2)\} = 2.287 - 0.244x_1 - 1.004x_2 - 0.740x_3$$

$$\text{logit}\{\text{Pr}(1 \text{ or } 2 \text{ or } 3)\} = 3.089 - 0.244x_1 - 1.004x_2 - 0.740x_3$$

回帰式から強度ごとの確率を求めた。最も確率の高い強度と実際の強度を比較すると、正解率はリポジトリが66.8%、補足資料は64.1%であった。

5. 考察

5.1 分野別データ共有ポリシーの状況と展望

グループ A の分子生物学・遺伝学などの 10 分野は全ての雑誌にポリシーがあり、定評のあるリポジトリも存在した。表 5 に示したりポジトリの生物データや結晶構造データはアクセッション番号によって識別することが可能であり、論文とのリンクが確立されている。また、これらのリポジトリで共有されたデータの相互運用性を向上させ⁶¹⁾、シームレスに利用できる環境を整えることによって多様なデータを統合して解析するオーミクス (omics) 研究も盛んに行われている⁶²⁾。以上のことから、さらなるデータの活用が期待できる。

グループ B のうち、地球科学、宇宙科学、社会科学はデータ共有が盛んであるが、リポジトリによるポリシーを掲載していたのは、それぞれ 7、6、4 誌であった。しかし、Zenk-Möltgen らの調査では、社

会学の 72.1%にポリシーが掲載されていた。また、社会学の実証研究については 1960 年代からデータ共有と二次分析や再現に関する議論が重ねられており⁶³⁾⁶⁴⁾、研究者によるオリジナルデータの再利用も行われている⁶⁵⁾ため、社会科学については分野を細分化した調査が必要だと考えられる。

農学は 7 誌、薬理学・毒物学は 6 誌がリポジトリによるポリシーを掲載していた。これらの分野では食品や医薬品など市民の健康や生命に関わる研究が行われているため、EMA（欧州医薬品庁）や FDA（米国食品医薬品庁）などを中心にデータ共有に関する検討が重ねられてきた⁶⁶⁾。また、臨床試験や生物データを活用するトランスレーショナルリサーチの隆盛⁶⁷⁾やデータ共有のニーズがある⁶⁸⁾ことから、今後ポリシーの掲載率や強度が上がる可能性が高いと考えられる。

グループ C のうち、物理学と精神医学・心理学は他の 4 分野に比べて IF の平均が高く、倫理に関する記述有りの雑誌数も多かったため、順序ロジスティック回帰分析の結果からポリシーの掲載率や強度が上がる可能性が高いと考えられる。一方、4 分野のうちリポジトリによるポリシーを掲載していたのは CS2 誌、数学 1 誌のみで、工学と経済学・ビジネスは 0 誌であった。この 4 分野はリポジトリも少なく（表 1）、データを用いない理論研究については共有のニーズがないと考えられる。しかし、*Journal of Statistical Software* のように、ソフトウェアやコードは GitHub (<https://github.com>) や SourceForge (<http://sourceforge.net>) で共有されてきた⁶⁹⁾。データ共有に関する検討や提言も行われている⁷⁰⁾⁷¹⁾ため、研究内容に応じたポリシーの策定やリポジトリの整備が必要であると考えられる。

5.2 雑誌の特徴とデータ共有ポリシー

(1)倫理や COI に関する記述がある雑誌はポリシーの掲載率と強度が高い

研究不正への対策はデータ共有の目的の一つとされることが多いが、本調査によって倫理や COI に関する記述がポリシーの掲載率や強度と関連があることが統計的に示された。直接的に関連を示す例として、*Neurology* は産業界がスポンサーの研究はデータを全て公開するよう求めていた。

(2)商業出版社はポリシーの掲載率と強度が高い

Piwovar らや Stodden らの先行研究とは逆に、商業出版社の方が学協会よりもポリシーの掲載率や強度が高かった。ただし、自然科学分野の補足資料の強度は学協会の方が高かった。本調査ではその要因までは明らかにできないが、商業出版社は雑誌の信用度やブランドを維持するために研究不正対策としてデータ共有を強く要求しているという可能性が考えられる。証左の一例として、有力誌ほど論文の撤回が多く、また撤回の通知は不透明な場合が多いが、その理由は著者や雑誌が体面を保つ、あるいは名誉毀損訴訟を避けるためではないかと推測されている⁷²⁾。

(3)OA 誌は全てポリシーを掲載していたが、OA は掲載率や強度が高いとはいえない

OA の 15 誌は全てポリシーを掲載していたが、OA 誌および制約つき OA 誌と非 OA 誌を比較すると、リポジトリ・補足資料によるポリシーの掲載率、ならびにリポジトリによるポリシーの強度に差はみられず、補足資料は非 OA 誌の強度が高かった。データ共有の背景として OA の理念が述べられることが多いが、OA ステータスはポリシーの掲載率や強度と関連があるとはいえなかった。

(4)出版国とポリシーの強度には関連がみられる

調査対象誌が 13 誌以上の 4 か国について、リポジトリによるポリシーの掲載率、およびリポジトリ・補足資料を「必須」とする雑誌の割合は、全て英国、ドイツ、米国、オランダの順であり、強度は国による差もみられた。論文の撤回に関する調査によれば、ドイツや米国など伝統的な科学大国の主著者による撤回理由は、データの改竄や捏造といった研究不正が多い⁷³⁾。こうした状況との関連を明らかにする分析は今後の課題としたい。

(5)ポリシーの強度が高い雑誌は IF や GSM の中央値が高い傾向にある

先行研究と同様に、ポリシーの強度が高い雑誌ほど IF や GSM の中央値が高かった。ただし、生物・医学のリポジトリは「弱い」と「なし」の IF の中央値が同程度であり、全分野の IF・GSM の順位と強

度には相関がみられなかった。IF が低い雑誌を含めた調査は今後の課題としたい。さて、IF の高さや研究不正による論文の撤回率には高い正の相関があるとする研究⁷⁴⁾や、IF が高い雑誌ほど研究不正があった論文の撤回までの時間が短いとする調査⁷⁵⁾がある。データ公開によって追試や検証が可能になることが、IF が高い雑誌ほど強度が高いことの一因であると考えられる。

(6) 刊行開始年はポリシーの掲載率や強度との関連が低い

先行研究では調査年が新しいほどポリシーの掲載率が高かったが、刊行開始年は新しいほどや掲載率や強度が高いという程度であった。

なお、OA ステータス以外は分野による差が大きく、結果に影響を与えたと考えられた。そこで分野グループごとの分析を行ったところ、自然科学は学協会の雑誌の比率や補足資料の強度が高い傾向がみられた。分野の影響を完全に制御することは困難であるため、対象誌数を増やして分野ごとの分析を行うことを今後の課題としたい。

5.3 データ共有ポリシーに関する知見と活用

本研究から得られた知見に基づいて、今後のポリシー策定に向けて次の2点を提案する。

(1) 定形ポリシーの策定

Elsevier や Nature Publishing Group などには定形のポリシーがあり、それぞれ複数の雑誌が採用していた。ICMJE 勧告のように標準化されたポリシーが普及すれば、研究者は投稿先ごとに対応を変える必要がなくなり、出版者はポリシー策定のコストを削減できるため、双方にとって有益であると考えられる。

(2) 既存のディレクトリや外部リポジトリの活用

リポジトリによるポリシーを掲載する 131 誌のうち 7 誌はデータの公開先を指示していなかったが、データ共有の阻害要因の 1 位は「十分な時間がない」(53.6%) で、4 位は「データを公開する場所がない」(23.5%) という調査結果がある⁷⁶⁾。Retrovirology のように既存のディレクトリからリポジトリを選ぶよう指示すれば、研究者がリポジトリ選択にかかる時間や共有を諦めるケースを減らすことが可能であろう。

補足資料の公開は出版者にとって負担となるため、多くの雑誌はデータ量やファイル数の制限、課金といった対策を講じていた。The Journal of Neuroscience は一論文あたりのデータ量が急増したことから、2010 年 11 月以降補足資料を廃止した⁷⁷⁾。一方、Systematic Biology は Dryad にデータを登録して補足資料に DOI を記述するよう指示していた。同誌のように、外部リポジトリを活用する方法は出版者にとって便益が高いと考えられる。

6. おわりに

雑誌によるデータ共有ポリシーを分野横断で調査した結果、生物・医学分野はリポジトリによるポリシーの掲載率や強度が高い傾向にあり、定評のある生物データのリポジトリへの公開を求めていた。しかし、農学、薬理学・毒物学、精神医学・心理学の掲載誌数はそれぞれ 7 誌、6 誌、2 誌であり領域内で差が見られた。企業の利益保護や被験者のプライバシーといった課題に対応したポリシーの整備が必要だと考えられる。地球科学、宇宙科学、社会科学はデータ共有が活発な分野であるが、二次分析が中心であり、リポジトリによるポリシーの掲載率は中程度であった。こうしたデータを異分野でも容易に活用できるよう、標準的な引用方式を採用したポリシーの策定が必要だと考えられる。そしてコンピュータサイエンス、数学、工学、経済学・ビジネスは、研究内容によってデータが用いられているものの、リポジトリによるポリシーは 0~2 誌にしか掲載されていなかった。従って、研究領域やニーズに応じたポリシーの策定やリポジトリの構築が必要だと考えられる。

本研究独自の知見として、雑誌のデータ共有ポリシーには、研究不正に関わる項目が関連していることが明らかになった。まず、倫理や COI の開示に関する記述がある雑誌はポリシーの掲載率や強度が高

い傾向が示された。間接的に関連があると考えられる特徴として、商業出版社の雑誌はポリシーの掲載率や強度が高い傾向にあり、出版国によってポリシーの強度に差がみられた。順序ロジスティック回帰分析の結果においても、ポリシーの強度にはIFや倫理に関する記述の有無が関連していた。しかし、データ共有ポリシーが研究不正に対する有効な抑制力となっているかどうかは未知である。ポリシーが学術情報流通に与える影響の実証的研究を次なる課題としたい。

謝辞

本研究は、2013年度日本図書館情報学会の研究助成を受けて実施しました。また、筑波大学の小野寺夏生名誉教授、緑川信之教授、辻慶太准教授から統計分析や執筆について多数のご助言を賜りました。心よりお礼申し上げます。

注・引用文献

- 1) Kowalczyk, Stacy and Kalpana Shankar. "Data sharing in the sciences," *Annual Review of Information Science and Technology*. Vol. 45, Issue 1, 2011, p. 247-294.
- 2) Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press, 2015, 383p.
- 3) McCain, Katherine W. "Sharing digitized research-related information on the World Wide Web," *Journal of the American Society for Information Science*. Vol. 51, No. 14, 2000, p. 1321-1327.
- 4) Assante, Massimiliano, et al. "Science 2.0 repositories: time for a change in scholarly communication," *D-Lib Magazine*. Vol. 21, No. 1/2, 2015. <http://doi.org/10.1045/january2015-assante>, (accessed 2015-09-21).
- 5) Candela, Leonardo, et al. "Data journals: a survey," *Journal of the Association for Information Science and Technology*. Vol. 66, Issue 9, 2015, p. 1747-1762.
- 6) Houghton, John and Nicholas Gruen. *Open Research Data: Report to the Australian National Data Service (ANDS)*. 2014, 45p. <http://www.ands.org.au/resource/open-research-data-report.pdf>, (accessed 2015-09-21).
- 7) National Research Council. *Access to Research Data in the 21st Century: An Ongoing Dialogue Among Interested Parties: Report of a Workshop*. National Academy Press, 2002, 52p.
- 8) Ioannidis, John P. A. "How to make more published research true," *PLoS Medicine*. Vol. 11, No. 10, 2014, e1001747. <http://doi.org/10.1371/journal.pmed.1001747>, (accessed 2015-09-21).
- 9) The Royal Society. *Science as an Open Enterprise*. The Royal Society Science Policy Center, 2012, 104p. <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>, (accessed 2015-09-21).
- 10) Arzberger, P., et al. "Promoting access to public research data for scientific, economic, and social development," *Data Science Journal*. Vol. 3, 2004, p. 135-152. <http://doi.org/10.2481/dsj.3.135>, (accessed 2015-09-21).
- 11) Jones, Sarah. "Development in research funder data policy," *The International Journal of Digital Curation*. Vol. 7, Issue 1, 2012, p. 114-125. <http://doi.org/10.2218/ijdc.v7i1.219>, (accessed 2015-09-21).
- 12) *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD, 2007, 23p. <http://www.oecd.org/sti/sci-tech/38500813.pdf>, (accessed 2015-09-21).
- 13) European Commission. *Guidelines on Data Management in Horizon 2020*. Ver.1.0. 2013, 6p. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, (accessed 2015-09-21).
- 14) Edwards, Paul N., et al. "Science friction: data, metadata, and collaboration," *Social Studies of Science*. Vol. 41, No. 5, 2011, p. 667-690.
- 15) Procter, Rob, et al. "Adoption and use of Web 2.0 in scholarly communications," *Philosophical Transactions A*. Vol. 368, No. 1926, 2010, p. 4039-4056. <http://doi.org/10.1098/rsta.2010.0155>, (accessed 2015-09-21).

- 16) Treloar, Andrew. "The Research Data Alliance: globally co-ordinated action against barriers to data publishing and sharing," *Learned Publishing*. Vol. 27, 2014, S9-S13. <http://doi.org/10.1087/20140503>, (accessed 2015-09-21).
- 17) DataCite. <https://www.datacite.org/>, (accessed 2015-09-21).
- 18) Foreign & Commonwealth Office. "G8 Science Ministers Statement," GOV.UK. 2013-6-13. <https://www.gov.uk/government/news/g8-science-ministers-statement>, (accessed 2015-09-21).
- 19) Borgman, Christine L. "The conundrum of sharing research data," *Journal of the American Society for Information Science and Technology*. Vol. 63, No. 6, 2012, p. 1059-78.
- 20) Databib. <http://databib.org/about.php>, (accessed 2015-09-21).
- 21) Anagnostou, Paolo, et al. "When data sharing gets close to 100%: what human paleogenetics can teach the open science movement," *PLoS ONE*. Vol. 10, No. 3, 2015, e0121409. <http://doi.org/10.1371/journal.pone.0121409>, (accessed 2015-09-21).
- 22) Whitlock, Michael C. "Data archiving in ecology and evolution: best practices," *Trends in Ecology & Evolution*. Vol. 26, Issue 2, 2011, p. 61-65.
- 23) RECODE project consortium. *Policy Recommendations for Open Access to Research Data*. RECODE, 2014, 44p. <http://policy.recodeproject.eu/assets/recode-handbook.pdf>, (accessed 2015-09-21).
- 24) Hudson, Eichard L. and Nuala Moran, eds., *The Data Harvest: How Sharing Research Data Can Yield Knowledge, Jobs, and Growth*. RDA Europe, 2014, 40p. https://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf, (accessed 2015-09-21).
- 25) Pepe, Alberto, et al. "How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers," *PLoS ONE*. Vol. 9, No. 8, 2014, e104798. <http://doi.org/10.1371/journal.pone.0104798>, (accessed 2015-09-21).
- 26) Aydinoglu, Arsev Umur, et al. "Data management in astrobiology: challenges and opportunities for an interdisciplinary community," *Astrobiology*. Vol. 14, No. 6, 2014, p. 451-461. <http://doi.org/10.1089/ast.2013.1127>, (accessed 2015-09-21).
- 27) Pham-Kanter, Genevieve, et al. "Codifying collegiality: recent developments in data sharing policy in the life sciences," *PLoS ONE*. Vol. 9, No. 9, 2014, e108451. <http://doi.org/10.1371/journal.pone.0108451>, (accessed 2015-09-21).
- 28) Darby, R., et al. "Enabling scientific data sharing and re-use," *Proceedings of the IEEE International Conference on e-Science (e-Science 2012)*. 2012, 8p.
- 29) Kim, Youngseek and Jeffrey M. Stanton. "Institutional and individual factors affecting scientists' data-sharing behaviors: a multilevel analysis," *Journal of the Association for Information Science and Technology*. 2015, [early view] , (accessed 2015-09-21).
- 30) Enke, Neela, et al. "The user's view on biodiversity data sharing: investigating facts of acceptance and requirements to realize a sustainable use of research data," *Ecological Informatics*. Vol. 11, 2012, p.25-33.
- 31) Vines, Timothy H., et al. "Mandated data archiving greatly improves access to research data," *The FASEB Journal*. Vol. 27, No. 4, 2013, p. 1304-1308. <http://doi.org/10.1096/fj.12-218164>, (accessed 2015-09-21).
- 32) 前掲 12) , p. 13-14.
- 33) *Recommended Practices for Online Supplemental Journal Article Materials*. National Information Standards Organization (NISO) and the National Federation of Advanced Information Services (NFAIS), 2013, 27p. http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf, (accessed 2015-09-21).

- 34) Piwowar, Heather A. and Wendy W. Chapman. "A review of journal policies for sharing research data," *Nature Proceedings*. 2008, 3p. <http://precedings.nature.com/documents/1700/version/1>, (accessed 2015-09-21).
- 35) Alsheikh-Ali, Alawi A., et al. "Public availability of published research data in high-impact journals," *PLoS ONE*. Vol. 6, No. 9, 2011, e24357. <http://doi.org/10.1371/journal.pone.0024357>, (accessed 2015-09-21).
- 36) Stodden, Victoria, et al. "Toward reproducible computational research: an empirical analysis of data and code policy adoption by journal," *PLoS ONE*. Vol. 8, No. 6, 2013, e67111. <http://doi.org/10.1371/journal.pone.0067111>, (accessed 2015-09-21).
- 37) Vlaeminck, Sven. "Data management in scholarly journals and possible roles for libraries: some insights from EDaWaX," *LIBER Quarterly*. Vol. 23, No. 1, 2013, p. 48-79. <http://liber.library.uu.nl/index.php/lq/article/view/8082>, (accessed 2015-09-21).
- 38) Zenk-Möltgen, Wolfgang and Greta Lepthien. "Data sharing in sociology journals," *Online Information Review*. Vol. 38, No. 6, 2014, p. 709-722.
- 39) 池内有為 「研究データ公開に関する学術雑誌のポリシー分析」 『三田図書館・情報学会研究大会発表論文集』 2013, p. 9-12.
- 40) Hey, Tony, et al. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. 1st edition. Microsoft Research, 2009, 284p.
- 41) 予備調査では IF の高い雑誌を選定したにも拘らず、外科学と社会学の各 10 誌にはりポジトリによるポリシーが掲載されていなかった。
- 42) "Journals List," Thomson Reuters. File updated 2013-9-3. <http://incites-help.isiknowledge.com/incitesLive/ESIGroup/overviewESI/esiJournalsList.html>, (accessed 2014-01-31).
- 43) Schofield, Paul N, et al. "Post-publication sharing of data and tools," *Nature*. Vol. 461, Issue 7261, 2009, p. 171-173.
- 44) Björk, Bo-Christer. "The hybrid model for open access publication of scholarly articles: a failed experiment?," *Journal of the American Society for Information Science and Technology*. Vol. 63, Issue 8, 2012, p. 1496-1504.
- 45) re3data.org (Registry of Research Data Repositories) <http://www.re3data.org>, (accessed 2015-09-21).
- 46) Pampel, Heinz, et al. "Making research data repositories visible: the re3data.org registry," *PLoS ONE*. Vol. 8, No. 11, 2013, e78080. <http://doi.org/10.1371/journal.pone.0078080>, (accessed 2015-09-21).
- 47) Molloy, Jennifer C. "The Open Knowledge Foundation: open data means better science," *PLoS Biology*. Vol. 9, No. 12, 2011, e1001195. <http://doi.org/10.1371/journal.pbio.1001195>, (accessed 2015-09-21).
- 48) Boulton, Geoffrey, et al. "Science as a public enterprise: the case for open data," *The Lancet*. Vol. 377, No. 9778, 2011, p. 1633-1635. [http://doi.org/10.1016/S0140-6736\(11\)60647-8](http://doi.org/10.1016/S0140-6736(11)60647-8), (accessed 2015-09-21).
- 49) Howard, Alex. *Data for the Public Good: How Data Can Help Citizens and Government*. O'reilly, 2012, 15p.
- 50) Open Knowledge. "Place Overview [2014]," Global Open Data Index. <http://index.okfn.org/place/>, (accessed 2015-09-21).
- 51) Google Scholar Metrics. <http://scholar.google.co.jp/intl/en/scholar/metrics.html>, (accessed 2014-4-8).
- 52) 2015 年に re3data.org と Databib は DataCite に統合された。
- 53) Day, Michael. "Toward distributed infrastructures for digital preservation: the roles of collaboration and trust," *The International Journal of Digital Curation*. Vol. 3, No. 1, 2008, p. 15-28. <http://doi.org/10.2218/ijdc.v3i1.39>, (accessed 2015-09-21).
- 54) TRAC: The Trustworthy Repositories Audit & Certification: Criteria and Checklist. Center for Research Libraries. https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf, (accessed 2015-09-21).
- 55) 前掲 2)

- 56) Perrino, Tatiana, et al. "Advancing science through collaborative data sharing and synthesis: NIMH collaborative data synthesis for adolescent depression trials study team including," *Perspectives on Psychological Science*. Vol. 8, No. 4, 2013, p. 433-444. <http://doi.org/10.1177/1745691613491579>, (accessed 2015-09-21).
- 57) CONSORT. <http://www.consort-statement.org>, (accessed 2015-09-21).
- 58) "Guidelines," The Committee on Publication Ethics (COPE). <http://publicationethics.org/resources/guidelines>, (accessed 2015-09-21).
- 59) "Recommendations," International Committee of Medical Journal Editors (ICMJE). <http://www.icmje.org/recommendations/>, (accessed 2015-09-21).
- 60) 学協会 (7 誌) と国際機関 (9 誌) は本部の所在地としたが, 2 機関は不明であったため Ulrichsweb.com の出版国のままとした。
- 61) 前掲 14)
- 62) Field, Dawn, et al. "Omics data sharing," *Science*. Vol. 326, No. 5950, 2009, p. 234-236.
- 63) Hyman, Herbert H. *Secondary Analysis of Sample Surveys: Principles, Procedures, and Potentialities*. Wiley, 1972, 347p.
- 64) Fienberg, Stephen E., et al. *Sharing Research Data*. National Academies Press, 1985, 284p.
- 65) King, Gary. "Replication, replication," *PS: Political Science and Politics*. Vol. 28, No. 3, 1995, p. 444-452.
- 66) Martinez, Marian Garcia, et al. "Risk-based approaches to food safety regulation: what role for co-regulation?," *Journal of Risk Research*. Vol. 16, No. 9, 2013, p. 1101-1121.
- 67) Robinson, Peter N. "Genomic data sharing for translational research and diagnostics," *Genome Medicine*. Vol. 6, Issue 9, 2014, [78]. <http://doi.org/10.1186/s13073-014-0078-2>, (accessed 2015-09-21).
- 68) Bardyn, Tania P., et al. "Translational researchers' perceptions of data management practices and data curation needs: findings from a focus group in an academic health sciences library," *Journal of Web Librarianship*. Vol. 6, Issue 4, 2012, p. 274-278.
- 69) Kelty, Christopher M. *Two Bits: The Cultural Significance of Free Software*. Duke University Press, 2008, 378p.
- 70) Bakis, N., et al. "Towards distributed product data sharing environments: progress so far and future challenges," *Automation in Construction*. Vol. 16, Issue 5, 2007, p. 586-595.
- 71) Anderson, Richard G, et al. "The role of data/code archives in the future of economic research," *Journal of Economic Methodology*. Vol. 15, Issue 1, 2008, p. 99-119.
- 72) Corbyn, Zoë. "Misconduct is the main cause of life-sciences retractions," *Nature*. Vol. 490, Issue 7418, 2012, p. 21. <http://doi.org/10.1038/490021a>, (accessed 2015-09-21).
- 73) Fang, Ferric C., et al. "Misconduct accounts for the majority of retracted scientific publications," *Proceedings of the National Academy of Science of the United States of America*. Vol. 109, No. 42, 2012, p. 17208-17033. <http://doi.org/10.1073/pnas.1212247109>, (accessed 2015-09-21).
- 74) 前掲 73)
- 75) Steen, R. Grant, et al. "Why has the number of scientific retractions increased?," *PLoS ONE*. Vol. 8, No. 7, 2013, e68397. <http://doi.org/10.1371/journal.pone.0068397>, (accessed 2015-09-21).
- 76) Tenopir, Carol, et al. "Data sharing by scientists: practices and perceptions," *PLoS ONE*. Vol. 6, No. 6, 2011, e21101. <http://doi.org/10.1371/journal.pone.0021101>, (accessed 2015-09-21).
- 77) Maunsell, John. "Announcement regarding supplemental material," *The Journal of Neuroscience*. Vol. 30, No. 32, 2010, p. 10599-10600. <http://www.jneurosci.org/content/30/32/10599.full>, (accessed 2015-09-21).