

学術論文におけるデータ引用の現状

Japan Open Science Summit 2019

G1 研究活動の新たな常識としてのデータ引用の実現に向けて

2019年5月27日

クラリベイト・アナリティクス

学術情報事業部

プロダクト・マネージャー

野村紀匡

発表内容

はじめに

- ✓ 本発表の目的

方法

- ✓ 使用データ, 分析対象ジャーナルの選定, 分析の視点等, 本発表の分析方法

結果

- ✓ データを引用する論文の割合についての各種比較・分析結果

まとめ

はじめに

【本発表の目的】

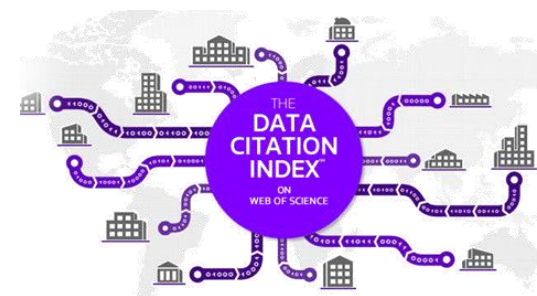
学術論文におけるデータ引用の現状を概観する。

【観点】

- データを引用する論文の割合は？
- 日本の論文がデータを引用する割合は？
- 研究分野間でデータを引用する傾向にどのような違いがあるか？
- ジャーナル間・出版者間ではどうだろうか？

分析に使用するデータ

- ❖ クラリベイト・アナリティクス社のWeb of Science Core Collection
 - ❑ Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI)
(2019年5月20日時点)
- ❖ 対象期間：出版年2014年から2018年
- ❖ 文献種類：「Article」及び「Review」
- ❖ 分野分類：同社Essential Science Indicators (ESI) の22分野分類
- ❖ 論文からデータ引用有・無：同社Data Citation Index (DCI)を参照
- ❖ 分野別データ引用率の分析結果を補足するため, Scholix JSON Dump (Bruzzo & Manghi, 2019) を参考データとして参照



Sandro La Bruzzo, & Paolo Manghi. (2019). OpenAIRE Scholix Explorer Service: Scholix JSON Dump [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.2674330>. (2019年5月5日公開版)

分析対象ジャーナルの選定

❖ 分析対象ジャーナル（学術雑誌）

- 先行研究（池内 & 逸村, 2016）にて調査対象として選定された, ESI22分野のJournal Impact Factor上位10誌（合計220誌）を, 本分析でも使用

池内有為; 逸村裕. 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. 日本図書館情報学会誌. 2016, vol. 62, no. 1, p. 20-37. https://doi.org/10.20651/jslis.62.1_20.

用語，分析の視点

【用語】

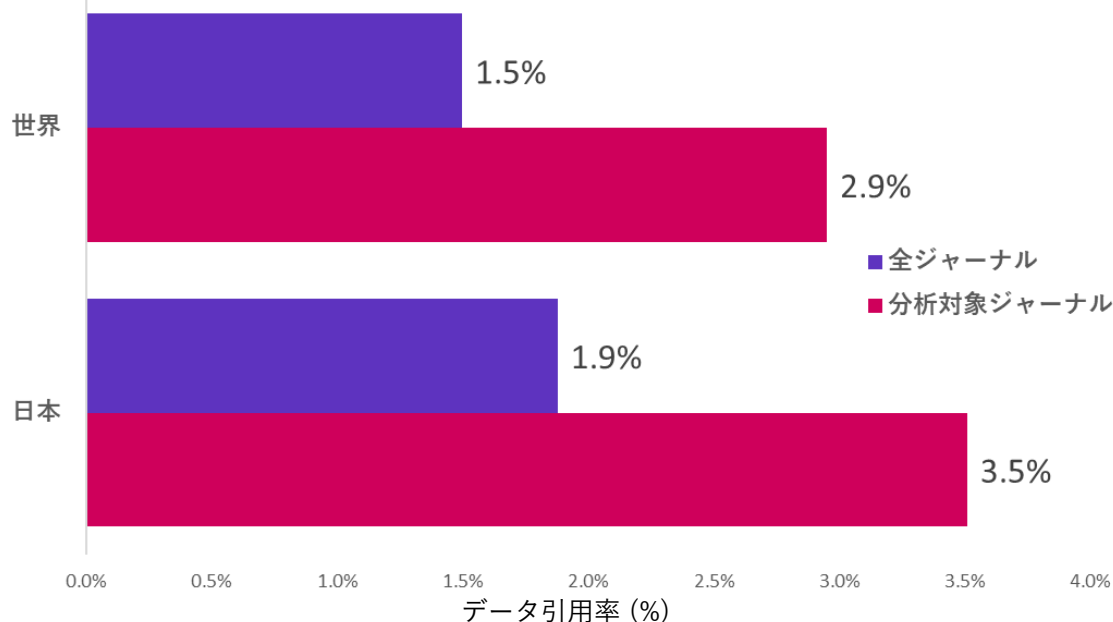
- 全ジャーナル：SCIE, SSCI 収録全ジャーナル（含分析対象ジャーナル）
- データ引用率：（データを引用する論文数）／（論文数）
- 「日本発論文」 又は 「日本の論文」：論文の著者に日本の機関に所属する研究者が一人以上含まれる論文

【分析の視点】

1. 全ジャーナルと分析対象ジャーナルにおけるデータ引用率の比較
 - ✓ 世界と日本発論文のデータ引用率についても比較
2. データ引用率の各種比較
 - ✓ 分野間
 - ✓ ジャーナル間
 - ✓ 出版者間

全ジャーナルと分析対象ジャーナルにおけるデータ引用率

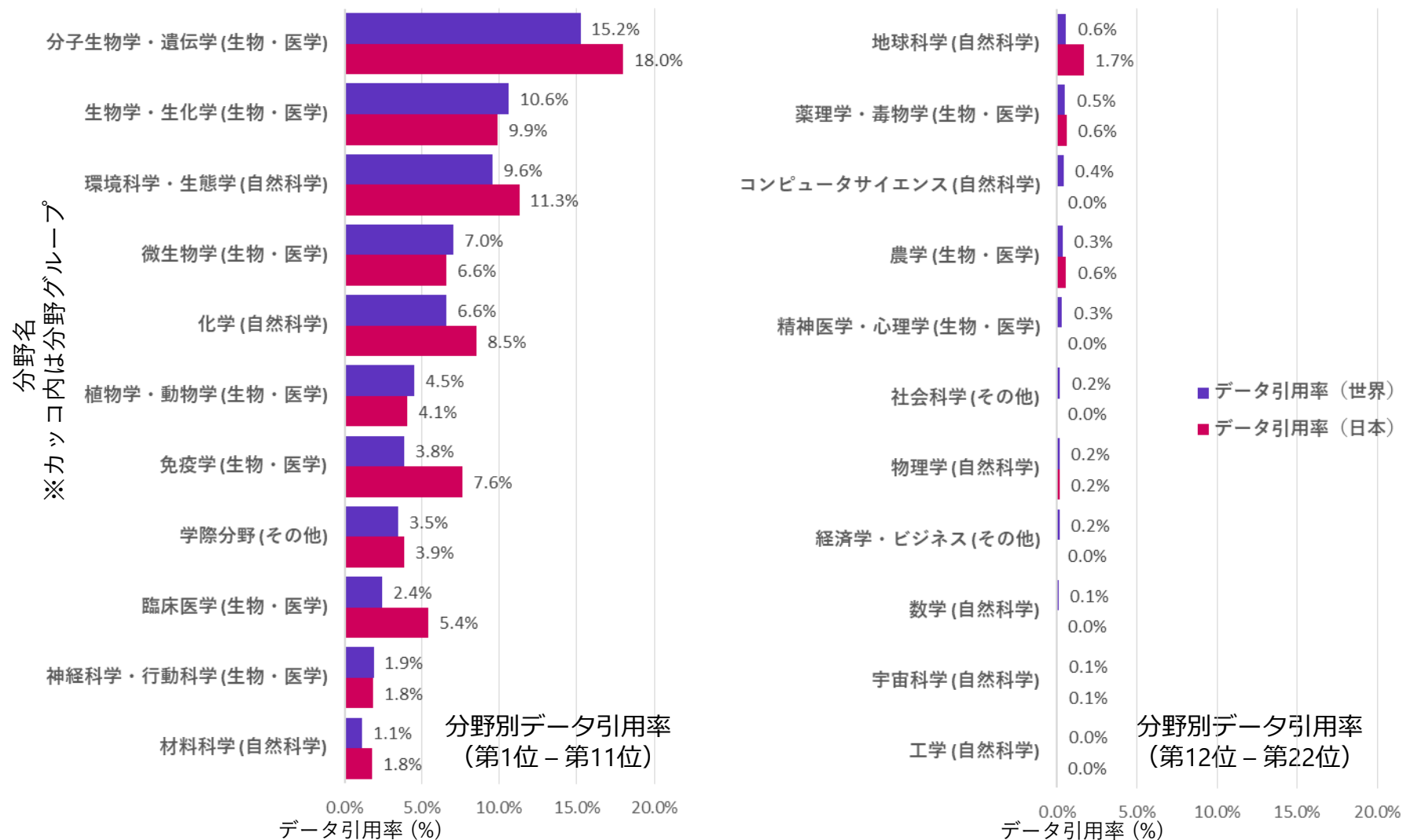
- ❖ 各分野のトップジャーナルにおいて、データ引用率が高い傾向がある。
- ❖ 日本発論文のデータ引用率は相対的に高い。



		論文数	データ引用論文 (%)
世界	全ジャーナル	7,975,413	119,330 (1.5%)
	分析対象ジャーナル	539,969	15,928 (2.9%)
日本	全ジャーナル	402,509	7,556 (1.9%)
	分析対象ジャーナル	36,797	1,291 (3.5%)

分野におけるデータ引用率（１）

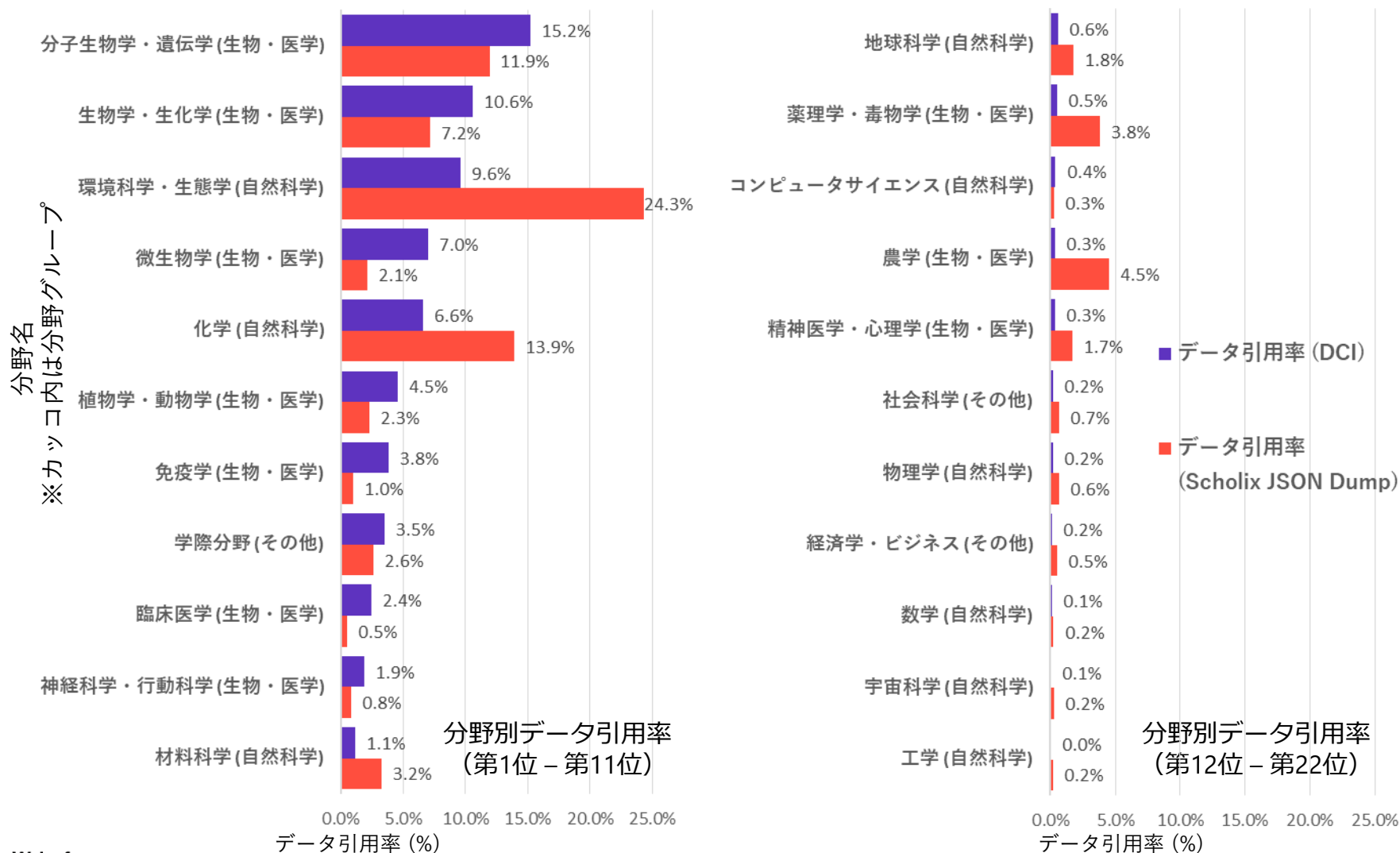
※分野グループは、（池内 & 逸村, 2016）
表6. 「分野グループ」を参照.



May 27th-28th 2019, Japan Open Science Summit 2019

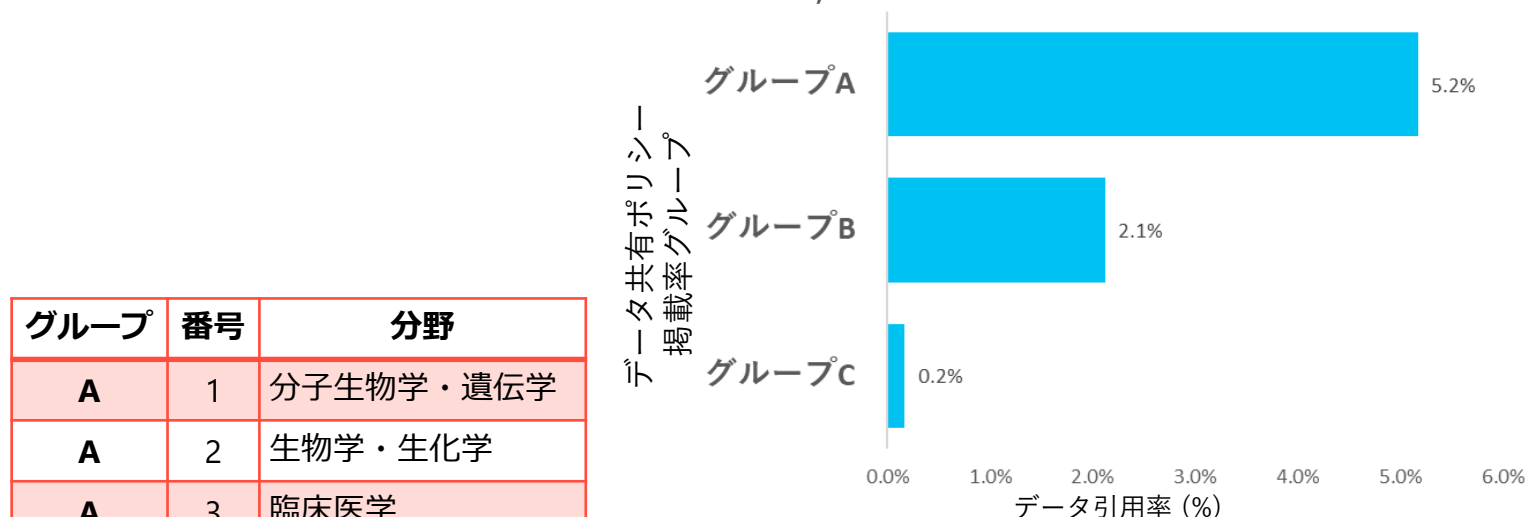
分野におけるデータ引用率（２）

※分野グループは、（池内 & 逸村, 2016）
表6. 「分野グループ」を参照.



分野におけるデータ引用率（3）

データ共有ポリシーの掲載率が高い分野ほど、データ引用率が高い。



グループ	番号	分野
A	1	分子生物学・遺伝学
A	2	生物学・生化学
A	3	臨床医学
A	4	微生物学
A	5	材料科学
A	6	植物学・動物学
A	7	化学
A	8	神経科学・行動科学
A	9	環境科学・生態学
A	10	免疫学

グループ	番号	分野
B	11	学際分野
B	12	農学
B	13	地球科学
B	14	薬理学・毒物学
B	15	宇宙科学
B	16	社会科学

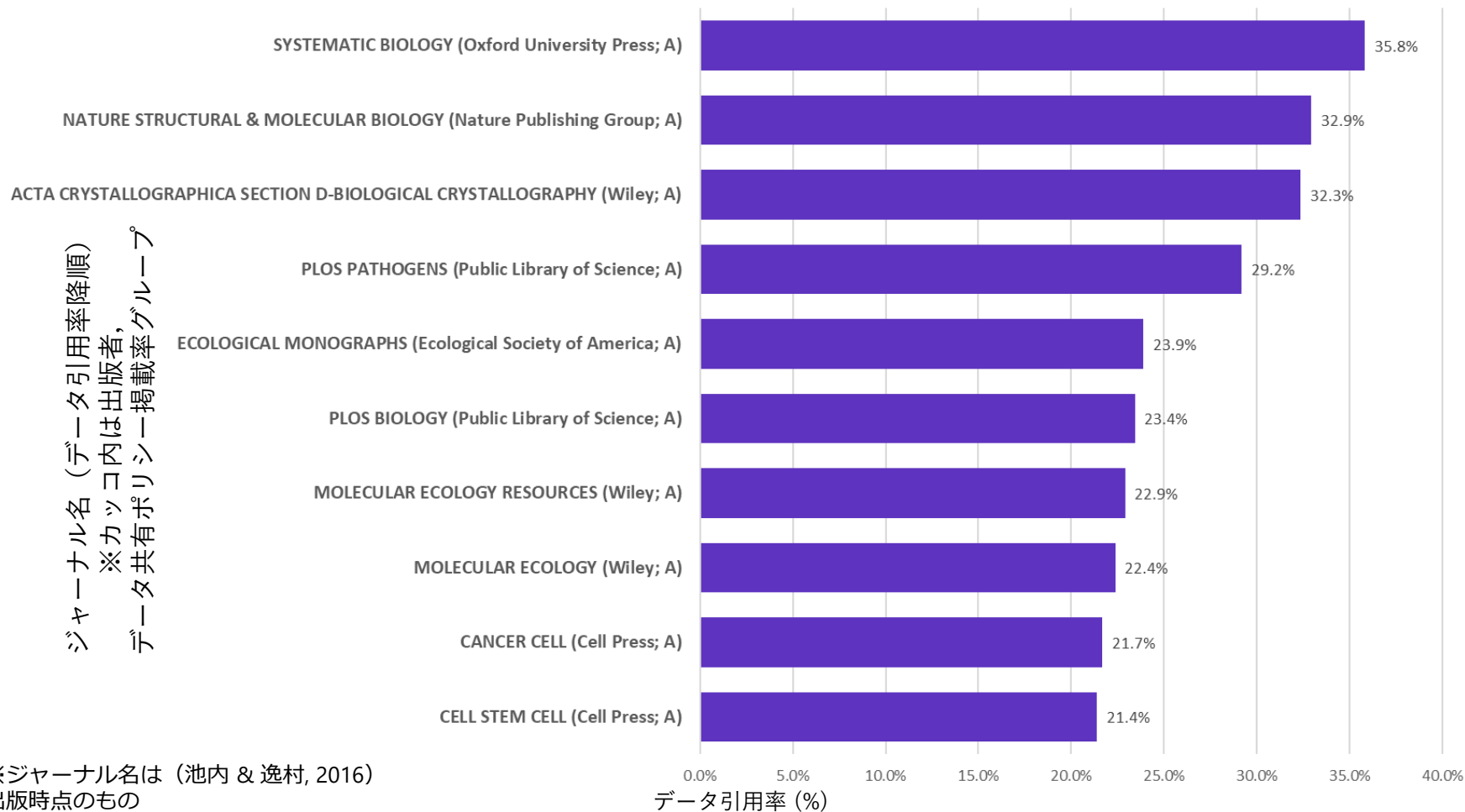
グループ	番号	分野
C	17	物理学
C	18	コンピュータサイエンス
C	19	精神医学・心理学
C	20	数学
C	21	工学
C	22	経済学・ビジネス

(池内 & 逸村, 2016) 表4. 「分野別データ共有ポリシーと雑誌の特徴」より、データ共有ポリシー掲載率によるグループ分けを抜粋。

May 27th-28th 2019, Japan Open Science Summit 2019

ジャーナルにおけるデータ引用率

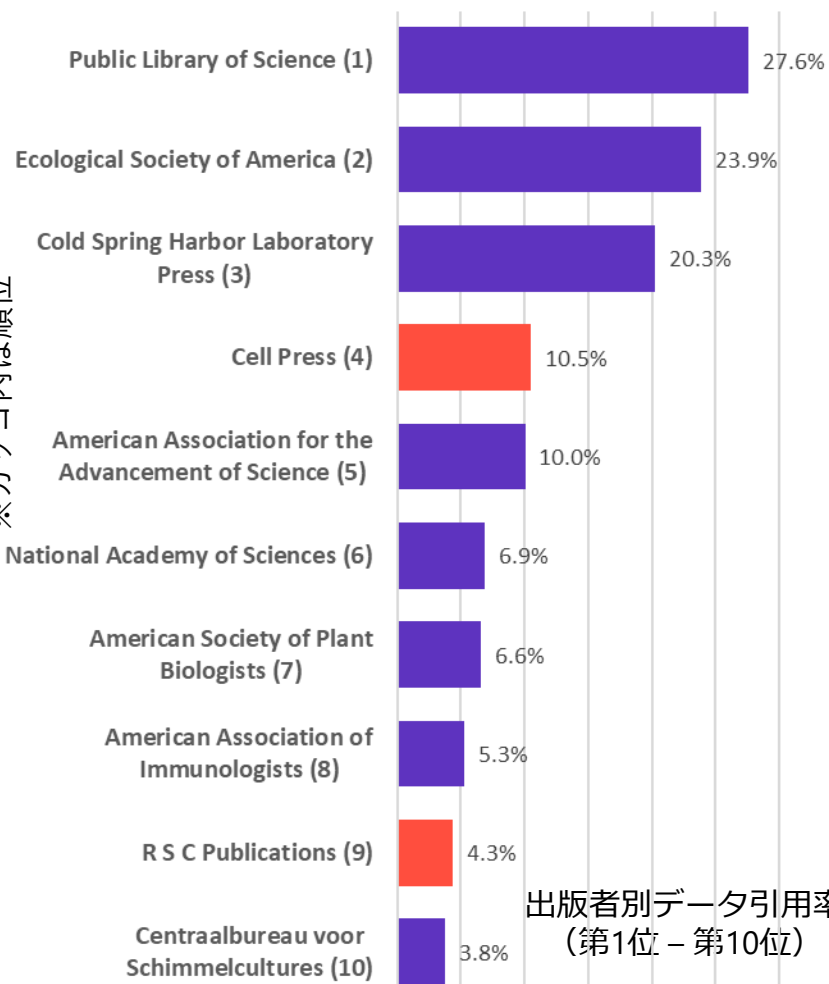
トップ30誌のうち、28誌がデータ共有ポリシーの掲載率が高い分野のジャーナル



※ジャーナル名は (池内 & 逸村, 2016)
出版時点のもの

出版者におけるデータ引用率（１）

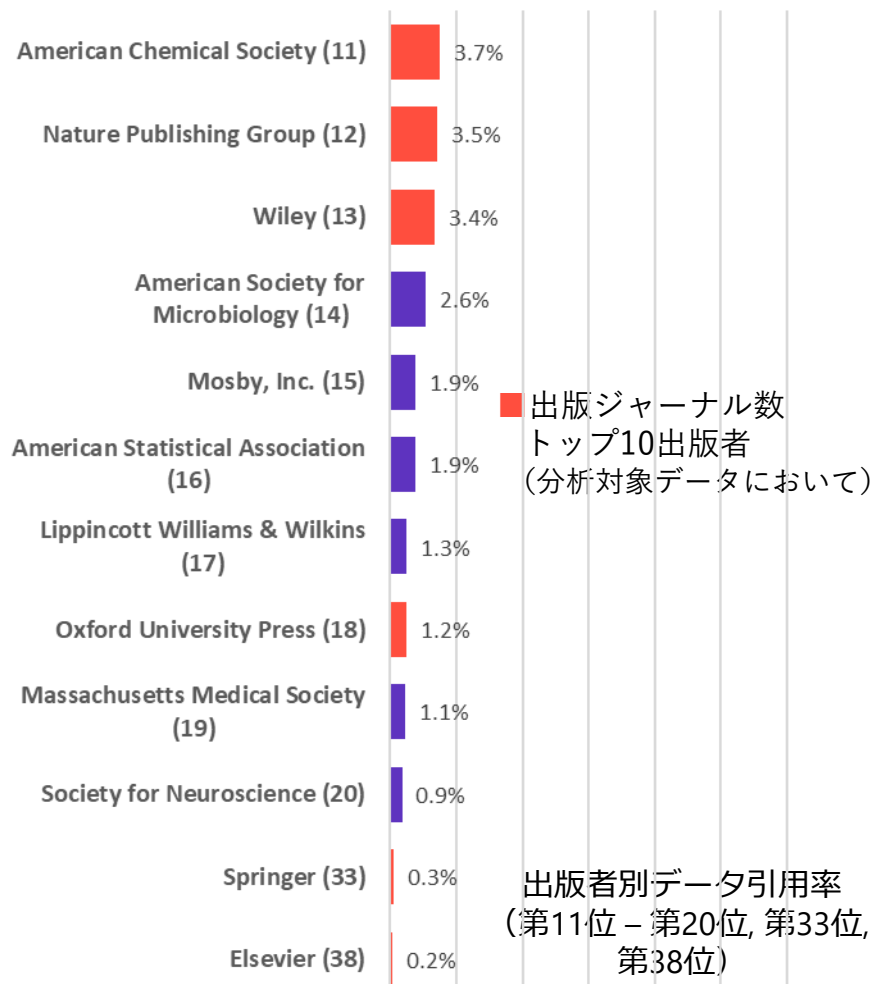
出版者名（データ引用率降順）
※カッコ内は順位



出版者別データ引用率
(第1位 – 第10位)

※出版者名は（池内 & 逸村, 2016）
出版時点のもの

0.0% 5.0% 10.0% 15.0% 20.0% 25.0% 30.0%
データ引用率 (%)



出版ジャーナル数
トップ10出版者
(分析対象データにおいて)

出版者別データ引用率
(第11位 – 第20位, 第33位,
第38位)

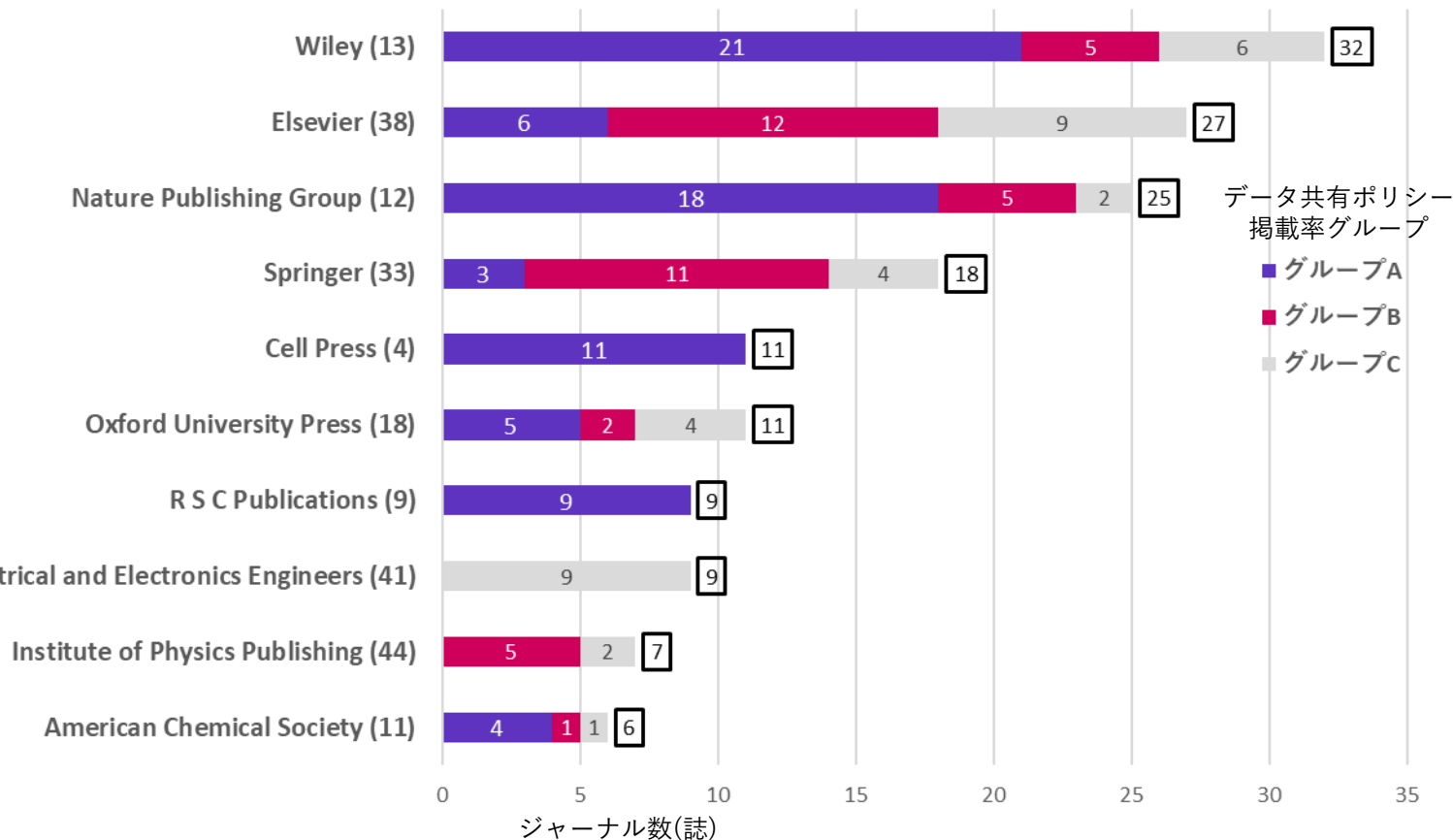
0.0% 5.0% 10.0% 15.0% 20.0% 25.0% 30.0%
データ引用率 (%)

May 27th-28th 2019, Japan Open Science Summit 2019

出版者におけるデータ引用率（２）

グループAのジャーナルをより多く発行する出版者は、データ引用率が高い。

出版者名（出版ジャーナル数降順）
※カッコ内はデータ引用率順位
※四角内の数字はジャーナル数計



※出版者名は（池内 & 逸村, 2016）
出版時点のもの

まとめ

- ❖ 各分野のトップジャーナルで発表される論文において、データ引用率が高い傾向がある。
- ❖ 日本発論文のデータ引用率は相対的に高い。
- ❖ データ引用率は分野によって異なり、データ共有ポリシー掲載率の高低による影響を受ける：
 - 掲載率が高い分野のデータ引用率はより高い。
 - 掲載率が高い分野のジャーナルのデータ引用率はより高い。
 - 掲載率が高い分野のジャーナルを発行する出版者のデータ引用率はより高い。