



プレプリントと研究データの 共有・公開の現状と課題： COVID-19が加速するオープンサイエンス

2021年2月17日

第13回政策研究レビューセミナー

文部科学省科学技術・学術政策研究所

科学技術予測センター

上席研究官 林 和弘

1. オープンサイエンスの潮流と政策，およびNISTEPの役割

2. 研究データ（と論文）の共有と公開に関する調査

- RM268, RM289（意識調査）

科学技術・イノベーション政策が注目する，オープンサイエンスの駆動要因としての，研究データの共有と公開に関するモニタリング

3. プレプリントの可能性と関連する調査

- DP168, DP187（計量書誌学からの調査）
- RM301（意識調査）

COVID-19によってにわかに注目が高まった，オープンサイエンスの駆動要因としての，プレプリントの共有と公開に関するモニタリング

4. 第6期科学技術・イノベーション基本計画に向けた 取り組み



1. オープンサイエンスの潮流と 政策とNISTEPの役割

1-1 国際的なオープンサイエンスの潮流

ICTの活用により知識をオープンにし、研究の加速化や新たな知識の創造などを促すオープンサイエンスの動きが活発化

EU:

- ヨーロピアン・オープン・サイエンス・クラウド (ESOC) が2018年に設立。
- Horizon Europeの枠組みでは研究データはFAIR原則に基づき、「できるだけオープンに、必要に応じてクローズド」にし、また研究データマネージメントを必須とする。

豪州:

- オーストラリア・リサーチ・データ・コモンズ (ARDC) を立ち上げ、研究データを中心とした情報基盤を構築中。

米国:

- 米国国立標準技術研究所(NIST)は研究データフレームワークの設立を検討中。
- OSTPはデータのアクセスや共有の在り方についてパブコメを実施。NIHがデータマネージメント&シェアリングポリシーを策定。

G7:

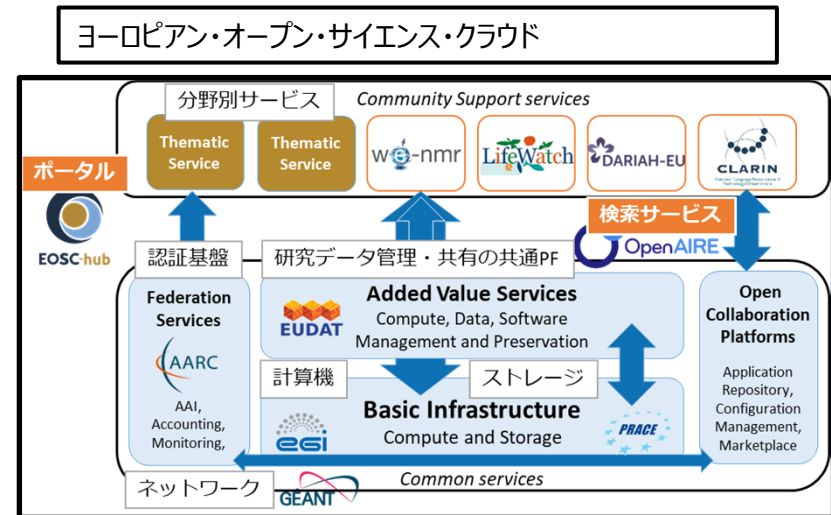
- オープンサイエンスWGを設置 (EUと日本が共同議長)。昨年6月、本年2月にWSを開催。

OECD:

- 研究データアクセスガイドラインを本年1月発行

UNESCO:

- オープンサイエンスに関する第1次勧告を昨年12月に発行



出典：国立情報学研究所 山地一禎教授 提供資料

諮問委員等
専門家として
貢献



1-2 日本のオープンサイエンス政策と施策例

2018

1 研究法人向け
データポリシーガイドライン

研究機関の公的資金による研究データの管理・活用ポリシー策定に関する
ガイドライン（仮称）（案）
2018.4.25

1本ガイドラインの目的
ICTの発展によって科学研究やその産業利用に新しい展開が生
じ加えて研究データの活用に注目が集まっている。我が国の研
究の創出のためには、重要な知的資産である公的資金による
データを、各研究開発法人のビジョン、ミッション等を踏まえた
についての考え方、方針に基づき適切に管理し、我が国の社会へ
成を含む産業利用等を意識しながら国際的対応を必要とする。このこと
を踏まえ、研究データの管理・活用に関する方針を定めることとする。

2017-

2 データマネジメントプラン
(DMP)
JST, AMED, NEDO

平成30年度より「データマネジメントプラン」提出を義務化。

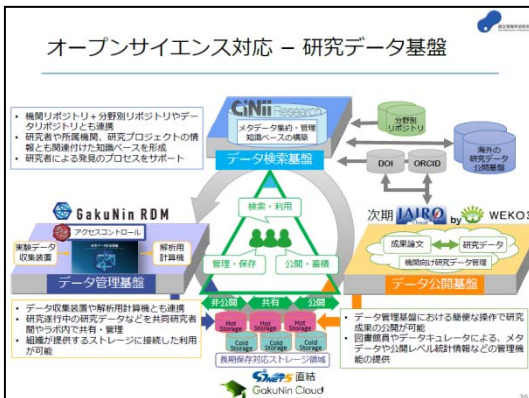
- データマネジメントプランは、どの研究開発課題で、どんなデータが産出され、誰がどこに保有しているのかを記載するものである。
- 併せて、基本的なデータ管理を進める上で必須となるデータサイエンティスト等を記載する。
- AMEDがデータの所在等を把握することにより、マネジメント機能または触媒機能を強化し、可能な範囲で異なる研究開発課題間での連携促進や二重研究開発の回避等に役立てる。
- 副次的な効果として、研究コミュニティでデータを適切に整理・体系化する文化が醸成されることにも期待する。

2019-

6 Moonshot
(内閣府)

Ongoing
(2020)

3 データインフラ
(管理・公開・検索)
NII



研究データマネジメント基盤
研究成果公開の主要メディア
(プラットフォーム)

統合・イノベーション戦略
Findable / Accessible Interoperable / Reusable

第5期科学技術基本計画

2019

4 データリポジトリガイドライン
(ドメイン別)

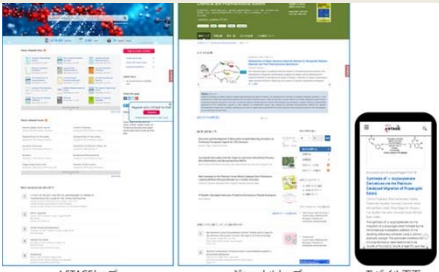
オープンサイエンスの推進の中で、データ共有のための環境が整えられ、国際的な認証機能が設定された。

このような国際的な環境に対応すべく、我が国では2019年12月に内閣府において有識者の検討会が立ち上げられ、我が国のオープンサイエンスに対する基本姿勢・基本方針等を示した報告書「我が国におけるオープンサイエンス推進のあり方について」が2019年3月に取りまとめられた。その際、2019年1月に閣議決定された「第5期科学技術基本計画」においても、我が国のオープンサイエンス推進では、学術情報委員会（第8期）となるデータを保存・公開することとを2019年2月に行っている。

さらに、2019年6月に閣議決定された「第5期科学技術基本計画」の策定に際して、我が国のオープンサイエンス推進のあり方に関する調査研究が実施され、その結果が「第5期科学技術基本計画」に反映されている。

2020

5 電子ジャーナルとの連携
(データ出版)
JST



7 データ共有等の
モニタリング
(NISTEP)

2019-



2. 研究データ（と論文）の共有と公開に関する調査

- ◆ RM268, RM289（意識調査）

科学技術・イノベーション政策が注目するオープンサイエンスの駆動要因としての研究データの共有と公開に関するモニタリング

NISTEP科学技術専門家ネットワークを対象としたデータ公開を中心としたオープンサイエンスの実態や課題を把握するための調査

- (1)データ公開と（比較として）OA論文の現状
- (2)データ公開の障壁
- (3)公開データの利用状況
- (4)DMPの作成状況（2018から）

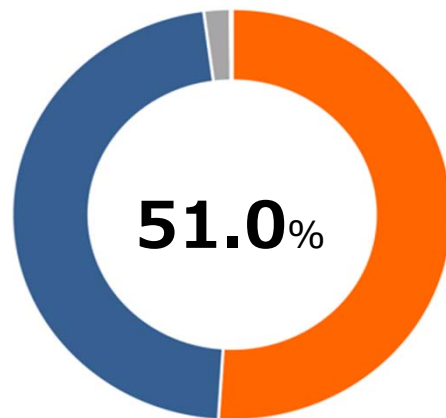
研究データ公開と論文のオープンアクセスに関する実態調査. 文部科学省科学技術・学術政策研究所, 2017, NISTEP RESEARCH MATERIAL No.268, <https://doi.org/10.15108/rm268>

研究データ公開と論文のオープンアクセスに関する実態調査. 文部科学省科学技術・学術政策研究所, 2020, NISTEP RESEARCH MATERIAL No.289, <https://doi.org/10.15108/rm289>

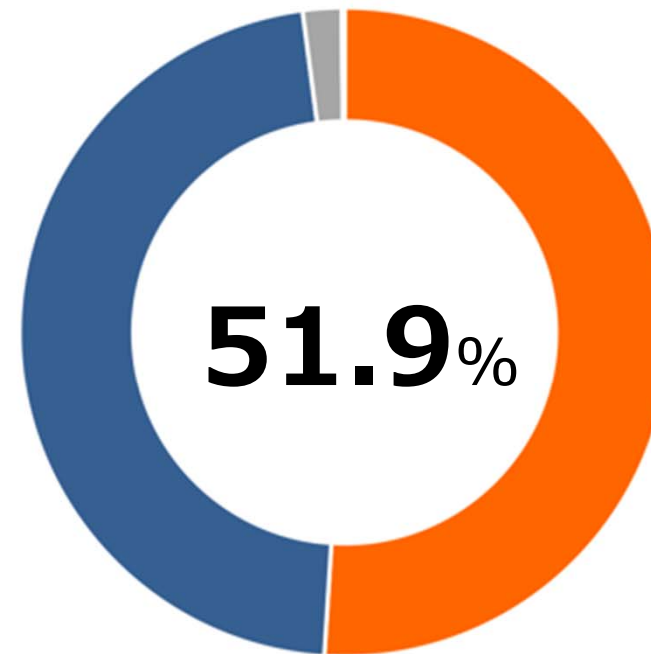


2-2 データ公開率

2016 (n=1,398)



2018 (n=1,516)



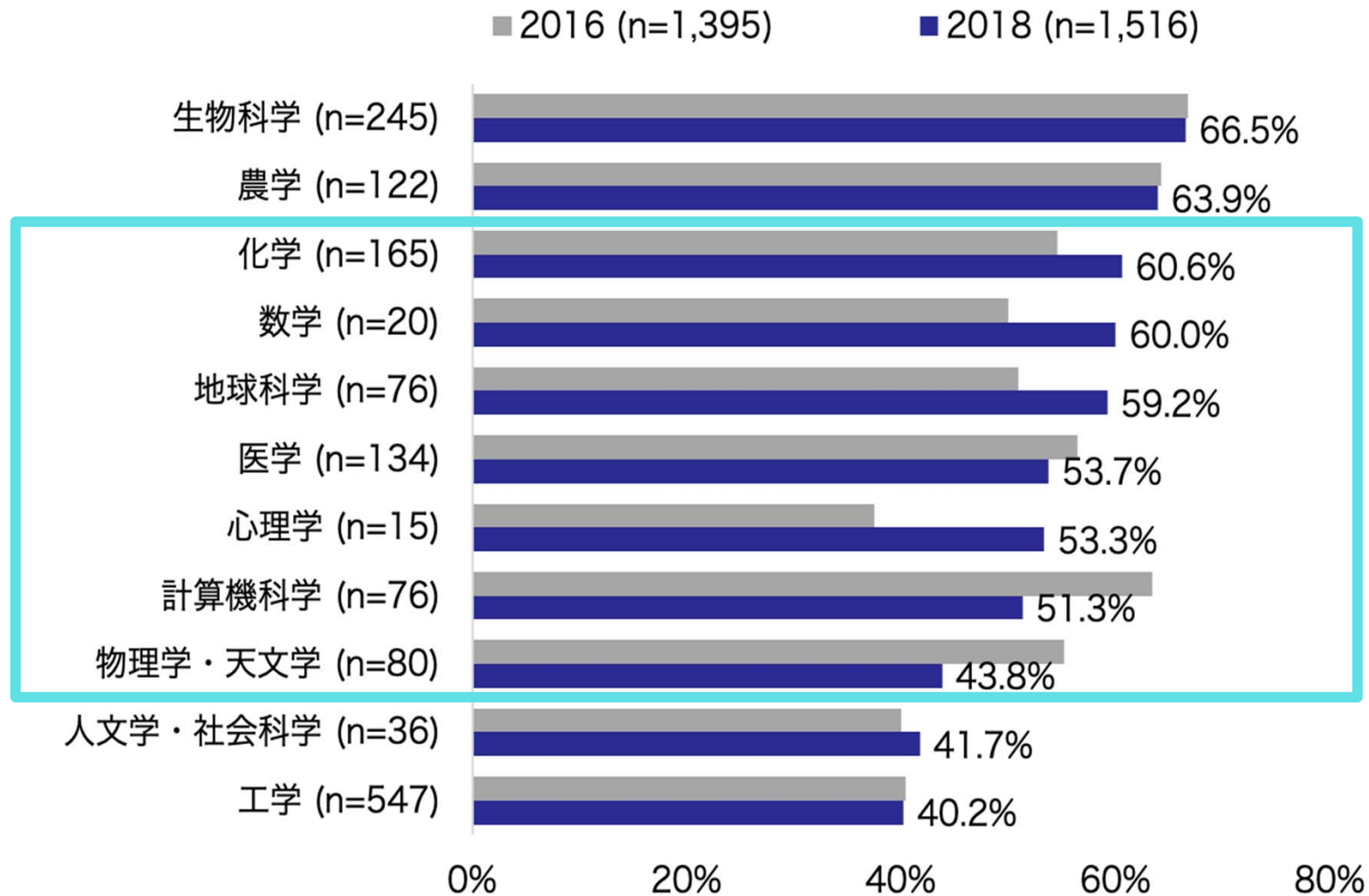
■ あり ■ なし ■ わからない ■ データは用いない

RM268

RM289

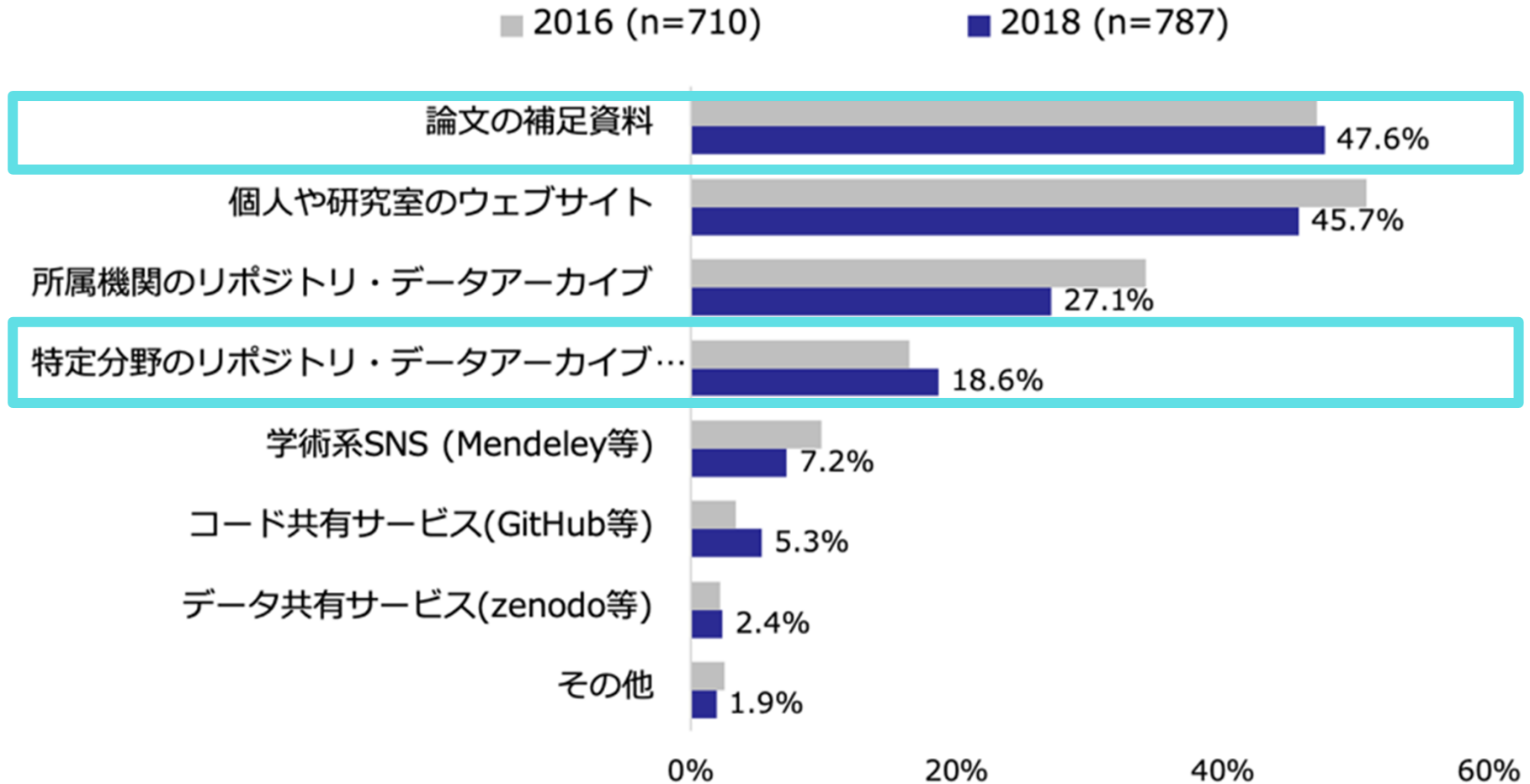
■ 統合イノベーション戦略をサポート

2-3 分野別データ公開率



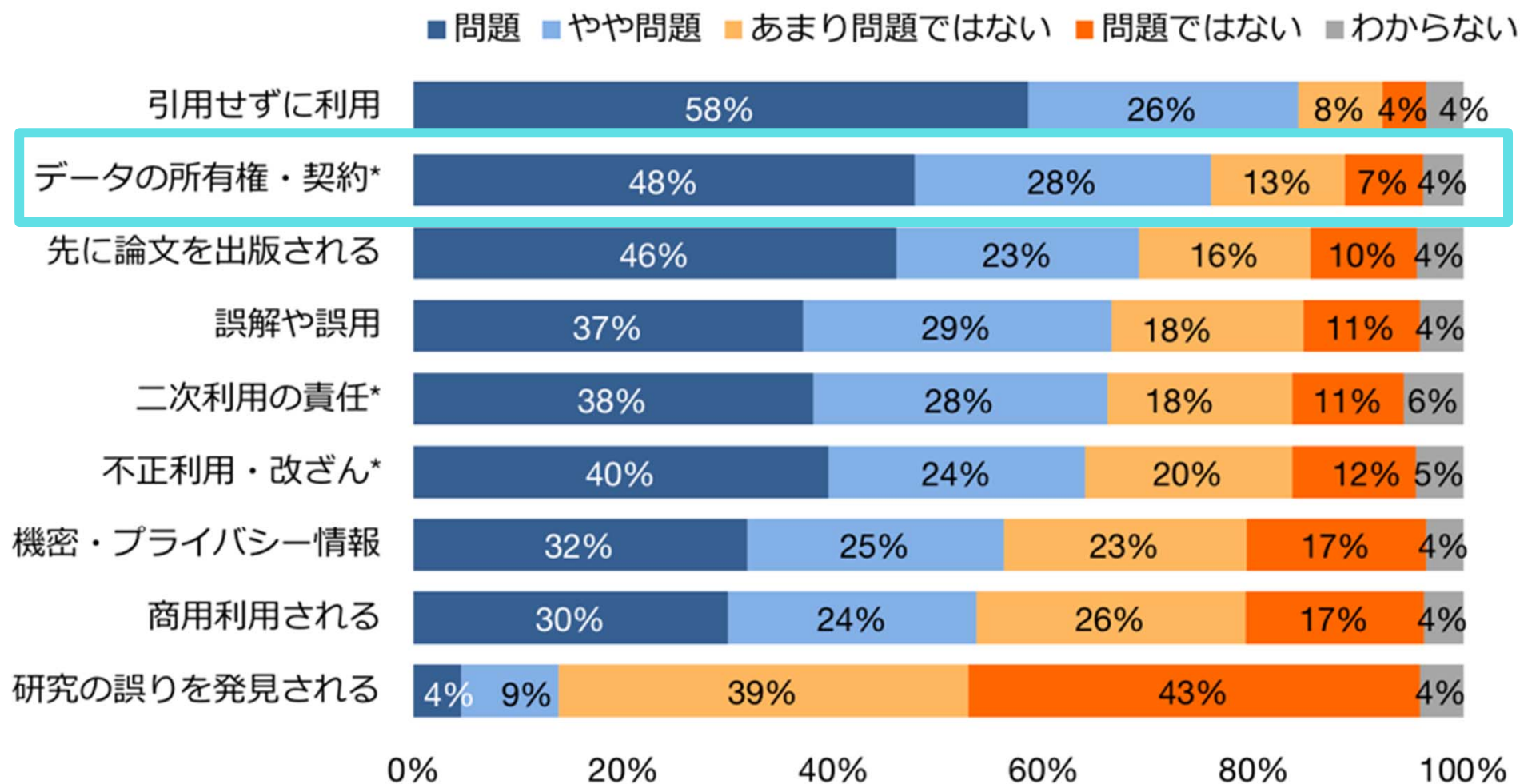
■ 上位は変わらないが中位に変動が見られた

2-4 データ公開の方法（複数回答）



■ 論文の補足資料が最上位に，特定分野のレポジトリも上昇

2-5 データ公開に対する懸念(2018)



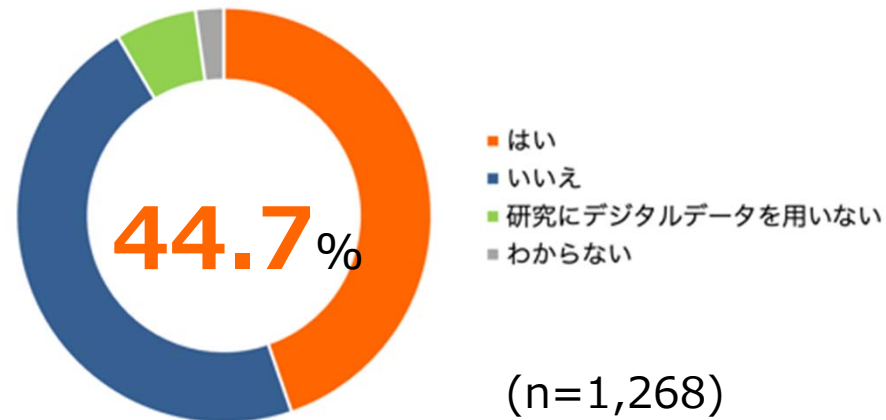
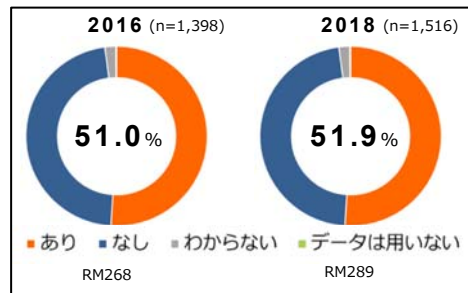
(n=1,513)

■ 日本学術会議の提言をサポート

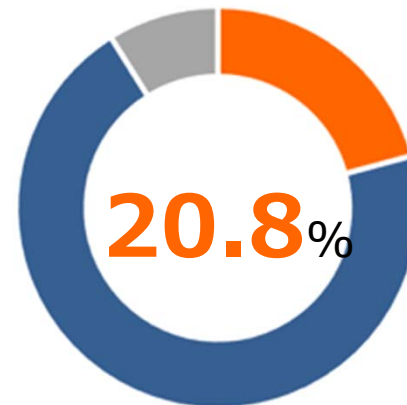
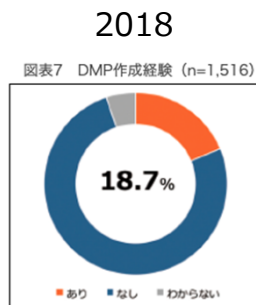
(1) データが中心的役割を果たす時代のルール作りの必要性

<http://www.scj.go.jp/ja/info/kohyo/kohyo-24-t291-1-abstract.html>

研究データの公開経験



データマネジメントプラン (DMP) の作成経験



■ 近日公開予定

■ はい ■ いいえ ■ わからない



3. プレプリントの可能性と 関連する調査

- ◆ DP186, DP187 (計量書誌学からの調査)
- ◆ RM301 (意識調査)

COVID-19によってにわかに注目が高まった
オープンサイエンスの駆動要因としての
プレプリントの共有と公開に関するモニタリング

■ ICTの進展によるデータ駆動型科学や情報流通の変容が進む中で顕在化した研究の新旧両フレーム

	従来の研究スタイル	新たな研究スタイル
研究の進め方	仮説・実証型	データ主導型
成果の公開方法	査読付き論文	プレプリント・研究データ
成果の価格	高価格化 (ジャーナル購読料の高騰)	無料・低価格
成果公開までのスピード	査読～公開までの長いタイムラグ	速やかに公開 (査読が無いため)
生まれる成果の量	少数の成果	大量の成果
公開される成果の信頼性	査読に基づく高い信頼性	質や信頼性のバラツキ増大 (誤った事実やフェイクの拡散の恐れ)
スタイルの持続性	高い持続性 (確立されたビジネスモデル)	不確定 (未確立のビジネスモデル)
主要国	欧米日等の先進国中心	中国や新興国の躍進
研究者のインセンティブ	ハイインパクトジャーナルでの発表による高い評価	研究実績の先取権確保
有効なシーンや分野	平常時に有効	非常時 (今回のコロナ対応等) に有効、技術進化の速い分野や査読に時間を有する分野に有効

内閣府研究データ基盤整備と国際展開WG(第13回)資料より抜粋

これらの変化自体はインターネット (Web) の発達とともに進行していたが、COVID-19によって大幅に加速

3-1 プレプリントとは

プレプリント

学術雑誌に投稿する予定の**査読・出版前の論文草稿**

プレプリントサーバー

プレプリントを掲載するインターネット（Web）サーバー

名称	創設年	2020年1月現在の 運営母体	分野	システム	DOI
arXiv	1991	コーネル大学	物理学に始まり、情報学、経済学等多分野に広がる	オリジナル	×
SSRN	1994	Elsevier	社会科学に始まり多分野に広がる	オリジナル (ColdFusion)	○
BioRxiv	2013	コールド・スプリング・ハーバー研究所	生命科学を中心とした分野	HighWirePress	○
ChemRxiv	2017	米国化学会、英国化学会、ドイツ化学会、日本化学会、中国化学会	化学を中心とした分野	figshare	○
MedRxiv	2019	米イェール大学、コールド・スプリングハーバー研究所、BMJ(British Medical Journal)	医学を中心とした分野	HighWirePress	○

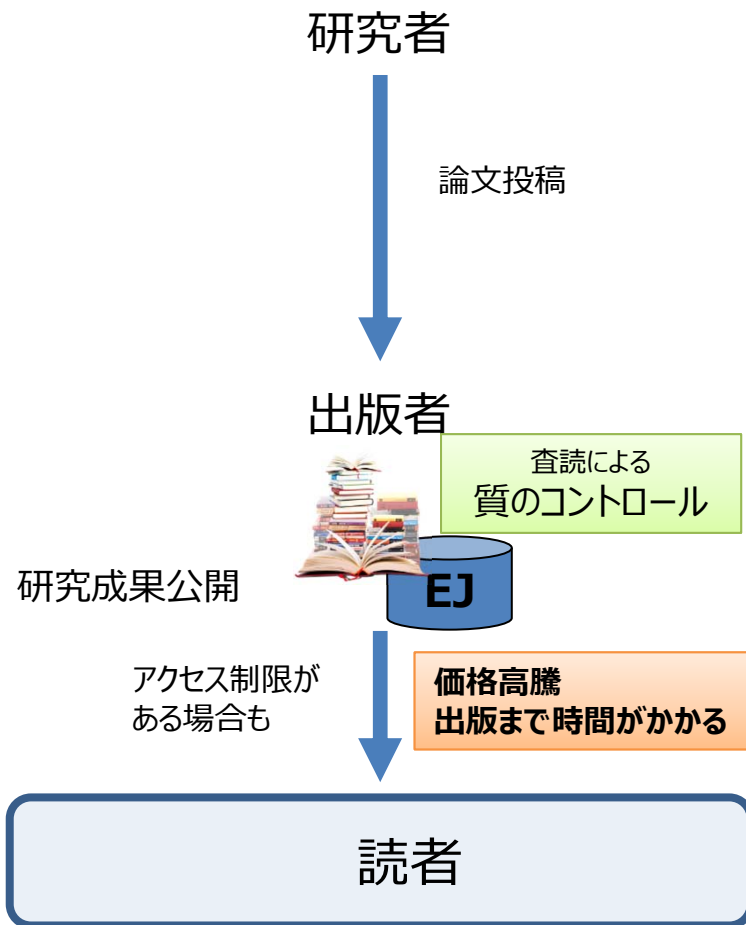
迅速な成果公開・共有手段として分野ごとに徐々に浸透中
学術情報流通を変革する研究データより身近なメディアとして注目

小柴等, 林和弘, 伊藤裕子. COVID-19 / SARS-CoV-2 関連のプレプリントを用いた研究動向の試行的分析.

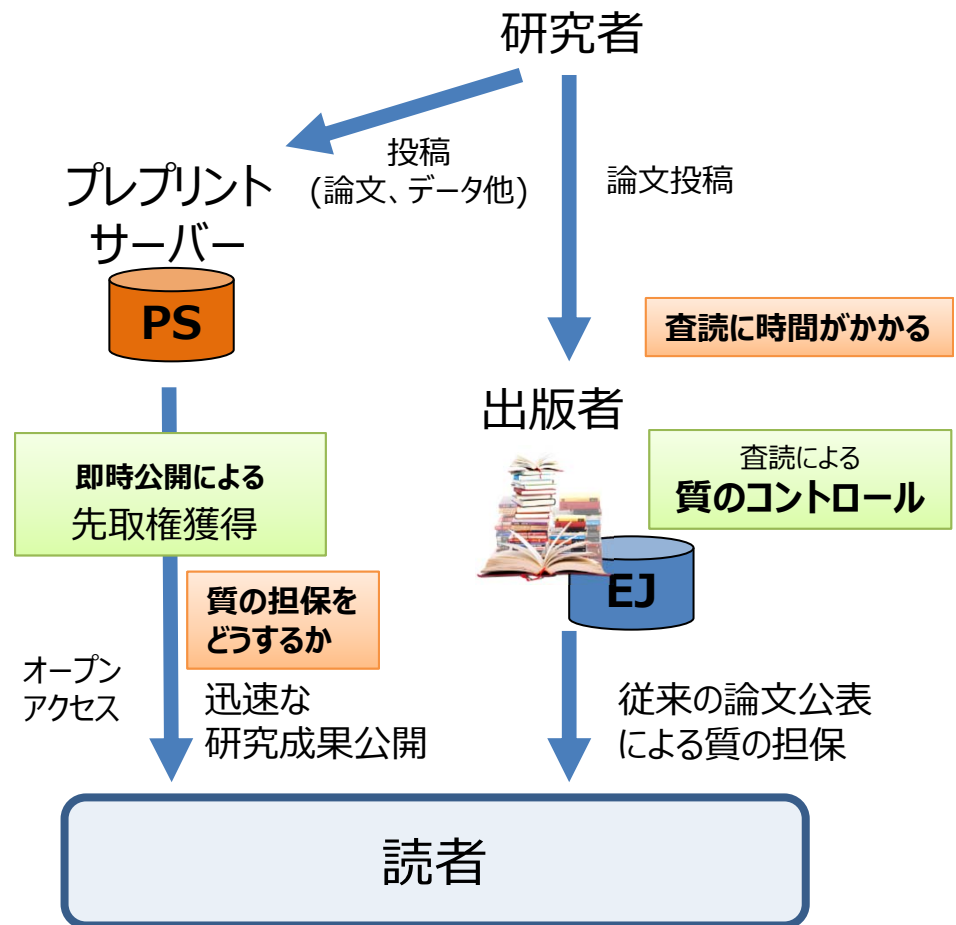
文部科学省科学技術・学術政策研究所, 2020, NISTEP DISCUSSION PAPER No. 186, 10p. <https://doi.org/10.15108/dp186>

3-2 プレプリントサーバ(PS)の活用による 研究成果のオープン化と先取権の確保

従来の仕組み



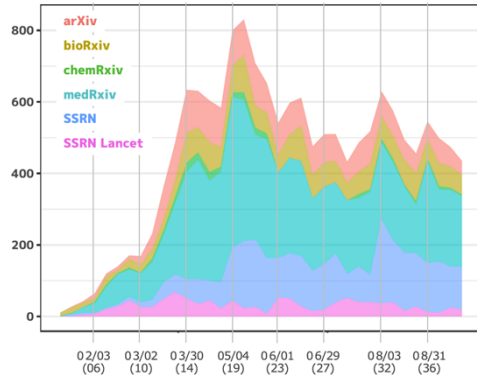
プレプリントサーバ(PS)の活用



林 和弘. 学術情報流通のオープン化がもたらすオープンサイエンスに向けた成果公開プロセスと共有の変革.
STI Horizon. 2017, Vol. 3, No. 3, p. 35-39. <https://doi.org/10.15108/stih.00092> 【一部改変】

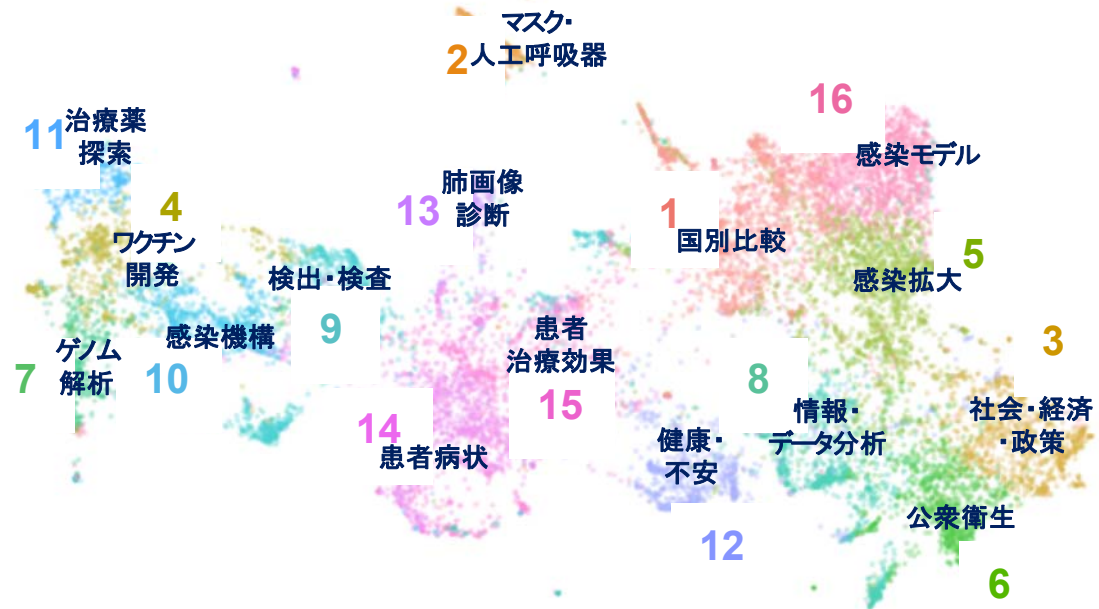
3-3 “COVID-19/SARS-CoV-2 関連のプレプリントを用いた研究動向の試行的分析” (DP186)

■ プレプリント解析で素早く見える研究動向 (2020.11.04 補遺 公表)



Week	arXiv	bioRxiv	chemRxiv	medRxiv	SSRN	Lancet
2020-01-20 (84)	0	9	1	0	1	0
2020-01-27 (85)	2	16	1	3	5	1
2020-02-03 (86)	4	13	0	16	5	5
2020-02-10 (87)	13	11	4	27	2	8
2020-02-17 (88)	7	24	6	59	4	20
2020-02-24 (89)	10	9	3	84	6	28
2020-03-02 (90)	10	24	6	76	10	45
2020-03-09 (91)	24	22	2	81	13	27
2020-03-16 (92)	42	27	10	106	20	26
2020-03-23 (93)	91	31	13	133	46	49
2020-03-30 (94)	105	47	18	198	50	68
2020-04-06 (95)	120	85	25	300	53	51
2020-04-13 (96)	182	70	20	334	70	35
2020-04-20 (97)	116	84	24	280	55	45
2020-04-27 (98)	111	54	14	309	73	23
2020-05-04 (99)	97	76	13	428	149	45
2020-05-11 (100)	100	104	21	395	188	23
2020-05-18 (101)	118	62	20	293	188	27
2020-05-25 (102)	83	59	15	333	155	8
2020-06-01 (103)	89	50	0	239	111	53
2020-06-08 (104)	88	66	0	267	127	50
2020-06-15 (105)	79	99	1	264	141	28
2020-06-22 (106)	80	66	0	262	113	15
2020-06-29 (107)	79	68	0	216	128	19
2020-07-06 (108)	75	60	0	199	136	40
2020-07-13 (109)	58	50	0	206	67	52
2020-07-20 (110)	81	63	11	191	100	40
2020-07-27 (111)	91	71	7	232	77	40
2020-08-03 (112)	69	64	11	213	236	37
2020-08-10 (113)	71	63	10	219	172	40
2020-08-17 (114)	56	70	4	186	162	16
2020-08-24 (115)	55	78	10	136	149	28
2020-08-31 (116)	48	55	8	286	136	13
2020-09-07 (117)	70	65	9	201	144	10
2020-09-14 (118)	54	60	6	214	115	25
2020-09-21 (119)	39	55	4	198	120	19

プレプリントサーバー別
週間登録数



自然言語処理によるトピックの可視化

原著論文，被引用数を使わずに研究動向の把握（ネットワーク分析）が可能に

詳細は次の発表にて

3-4 "arXivに着目したプレプリントの分析" (DP187)

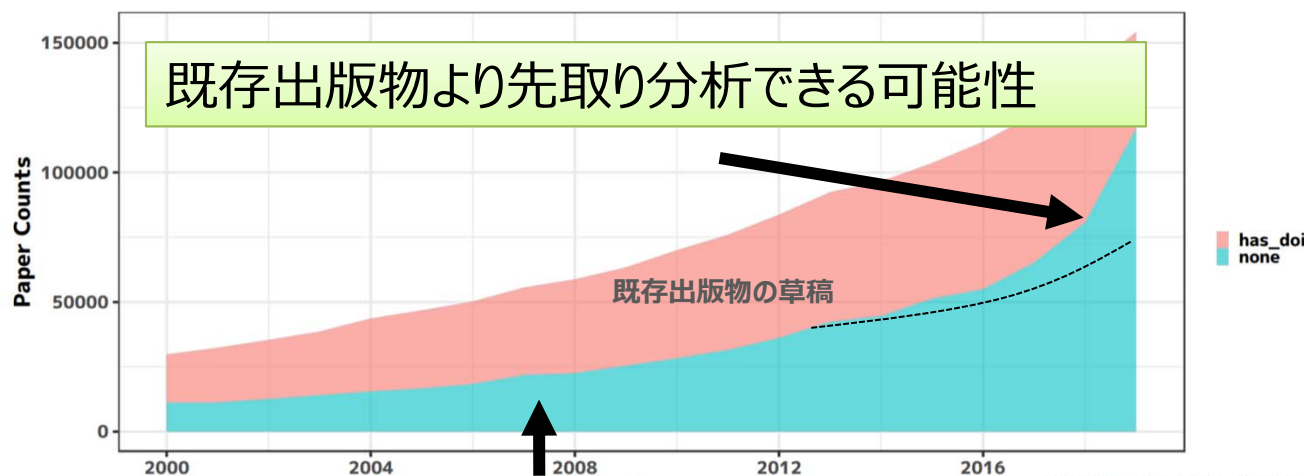
- arXiv 1991年より物理から始まり、昨今では、AI関連の投稿も多く、物理・数学・情報系で著名
 - ◆ 国際会議なども重視され、論文だけでは動向を追いつらいとされる情報系に有用

原著論文だけでは見えない分野の動向把握



2019年には年間投稿数が約15万件、累積で160万件を超える

既存出版物より先取り分析できる可能性



出版年ごとにDOIの有無（≒査読付きジャーナル等の既存の出版物に出ているもの）を見るとプレプリントならではの分析の可能性が見える

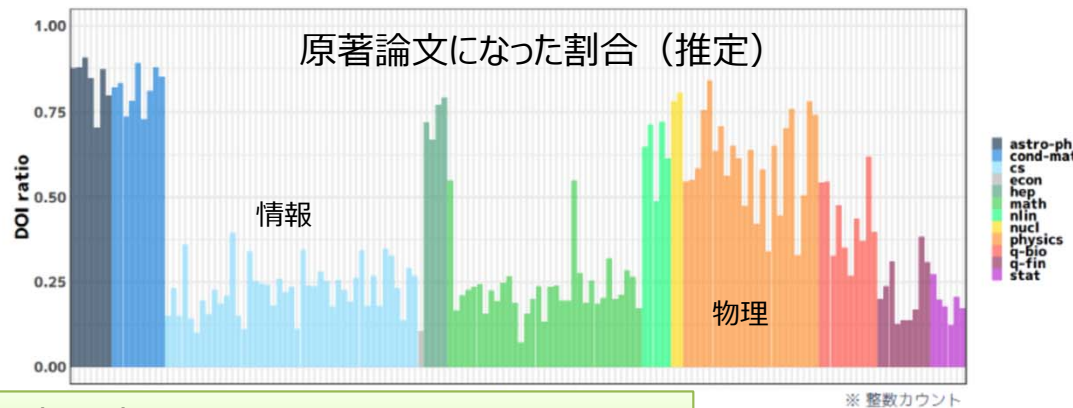
(点線は主観による参考)

出版バイアス等で除外された知見の可能性

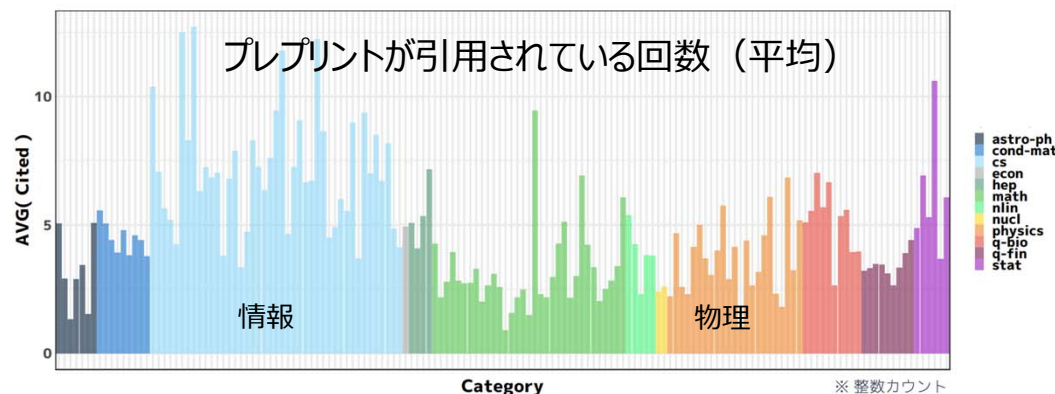
※累積ではありません

3-5 (プレプリントに依拠した研究分野の捕捉)

- ◆ 情報系で定性的に言われていたことをある程度定量的に裏打ち
 - 原著論文ではなく、国際会議のプロシーディングスを重要視
 - プレプリントを参考に研究を進展させ、プレプリントを引用してプレプリントを共有 (Deep Learning)
- ◆ COVID-19の前より、査読ジャーナルの問題は顕在化しており、分野依存ながら代替手段も存在していた (研究サイクルの速さにジャーナル査読が追いつかない)



プレプリントが引用される情報学



天文

図7 分野と被引用の関係性

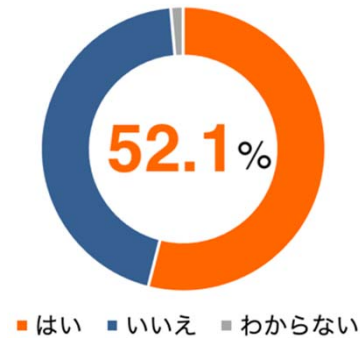
astro-ph	天文
cond-mat	材料
cs	情報
econ	計量経済
hep	高エネルギー物理
math	数学
nlin	システム科学
nucl	核
physics	物理
q-bio	生物
q-fin	金融
stat	統計

arXivが設定した153の小分野を独自に12分類

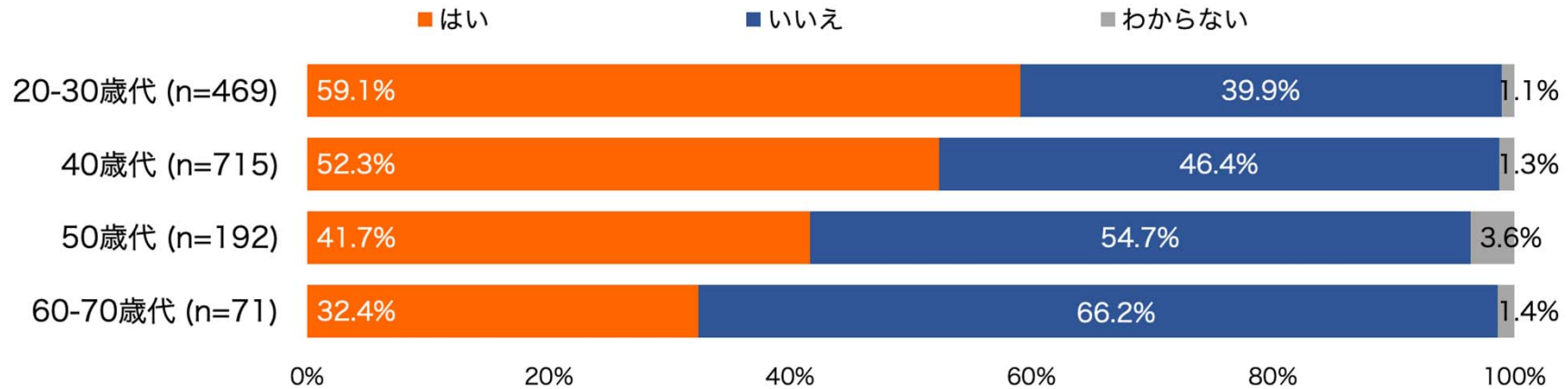
詳細は <https://doi.org/10.15108/dp187>

- 調査対象
 - 科学技術専門家ネットワーク：1,914名
 - 有効回答：1,448名（回答率75.7%）
- 調査期間
 - 2020年8月17日～8月31日（9月6日まで）

プレプリントの入手経験（全体）



年齢層別プレプリントの入手経験

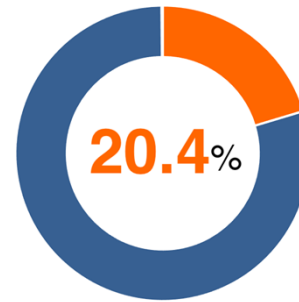


世代交代による変容を示唆

(n=1,447) ※不明(1名)を除く

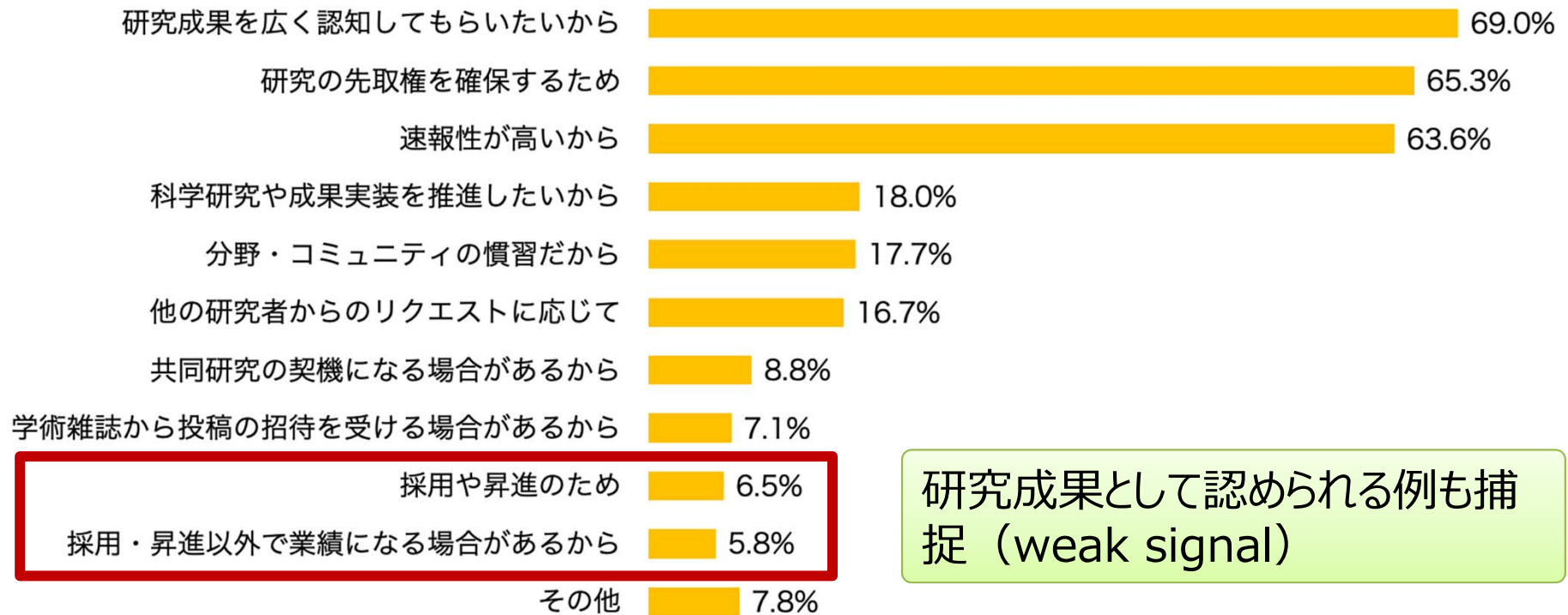
3-7 (プレプリントの公開と理由)

プレプリントの公開経験 (全体)



(n=1448)

■ はい ■ いいえ ■ わからない

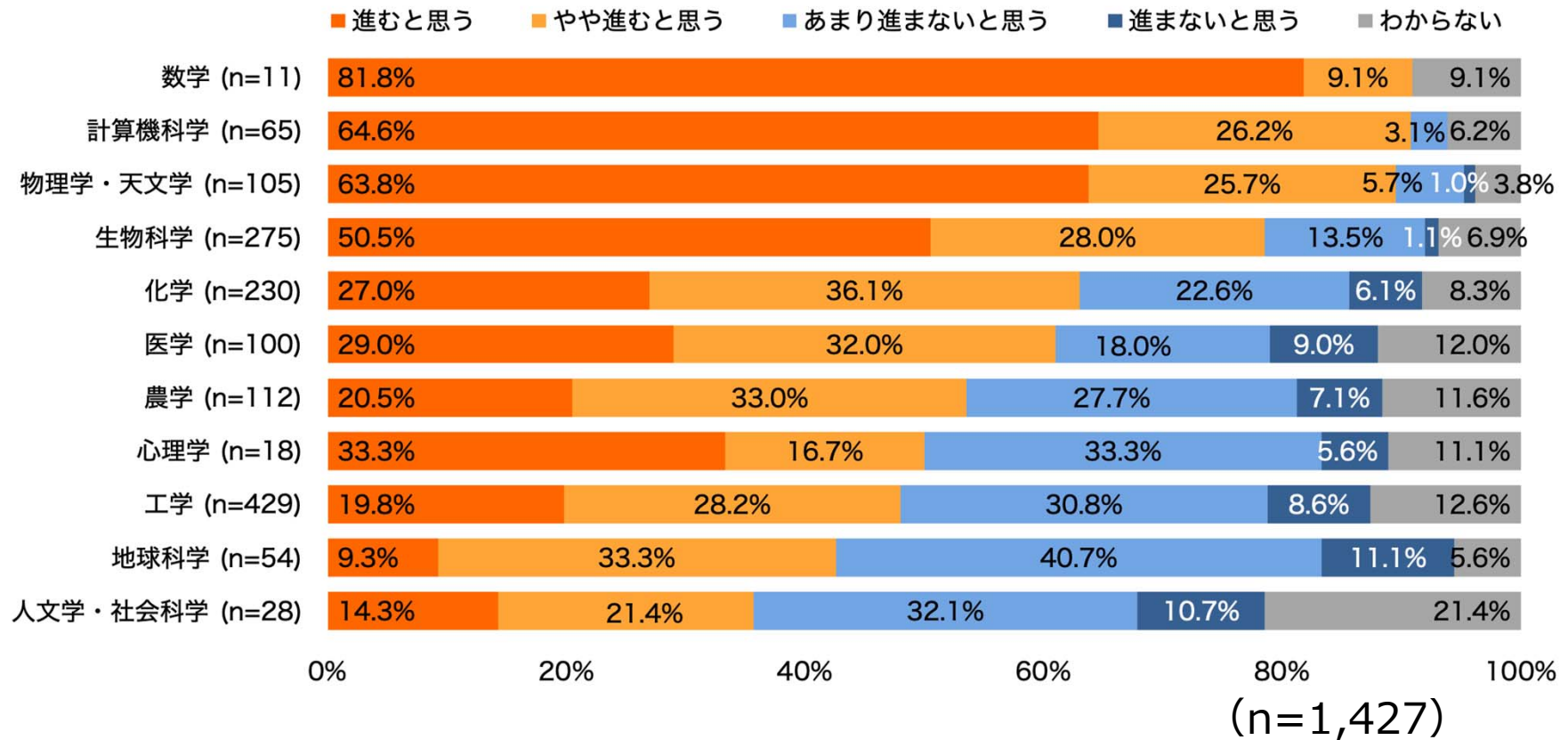


研究成果として認められる例も捕捉 (weak signal)

(n=294・複数回答)

3-8 (分野別のプレプリントの展望)

Q (自身に最も近い研究分野を選んだ後)
その分野では、今後プレプリントの利用が進むと思われますか。



多くの分野で進展を見込む割合が多い

詳細は <https://doi.org/10.15108/rm301>



4. 第6期科学技術・イノベーション基本計画に向けた取り組み

1. 「第6期科学技術・イノベーション基本計画」答申素案

✓ 2021年1月公開（2/10パブコメ終了）

2. 知のフロンティアを開拓しイノベーションの源泉となる研究力の強化

“また、社会全体のデジタル化の中で、我が国が世界に伍しながら、高付加価値でインパクトの高い研究を創出していくため、オープンサイエンスを含め、データ駆動型の研究の実施など、新しい研究の潮流を踏まえた研究システムを構築していく。”

(2) 新たな研究システムの構築（オープンサイエンスとデータ駆動型研究等の推進）

“社会全体のデジタル化や世界的なオープンサイエンスの潮流を捉えた研究そのもののDXを通じて、より付加価値の高い研究成果を創出し、我が国が存在感を発揮することを目指す。特に新型コロナウイルス感染症の研究においても、論文のオープンアクセス化やプレプリントの活用が更に拡大する中、研究プロセス全般で生まれるデータについて、戦略性を持って適切な共有と利活用を図るとともに、それによりインパクトの高い研究成果を創出していくための研究基盤の実現が求められる。”

【現状データ】（参考指標）

- ・ 国立研究開発法人における研究データポリシーの策定法人数：11 法人（2020 年 9 月）
- ・ 競争的研究費制度におけるデータマネジメントプラン（DMP）の導入済み府省・機関数：8 省・機関（2020 年 9 月）
- ・ 国内における機関リポジトリの構築数：811 個（2019 年度）
- ・ 研究データ公開の経験のある研究者割合：51.9%（2018 年度）
- ・ プレプリント公開の経験のある研究者割合：20.4%（2020 年度）
- ・ H P C I 提供可能資源量：年間 25 ペタ flops（2019 年度）
- ・ 研究設備・機器の共用化の割合：大学等における研究機器のうち相当程度の市場規模のある 10 機器（2012～2016 年度購入）について、競争的資金で購入したもののうち 9 割は研究者個人や研究室単位での利用に留まる。

ICTの活用による知識のオープン化，ならびに研究の加速化や新たな知識の創造などを促すオープンサイエンスの動きをモニタリング→指標化の検討

■ NISTEP科学技術専門科ネットワーク

- ◆ 科学技術予測センターが運営する産学官の研究者、技術者、マネージャーなど約2,000名の専門家集団
- ◆ 多分野かつ幅広い年齢層の回答者が含まれる
- ◆ 活用事例
 - 文部科学省戦略目標等における開発目標決定に寄与（研究動向の作成）
 - 東日本大震災後の科学者・技術者の意識に関する調査→平成24年度の科学技術白書に引用される
 - 研究施設・機器の共用化等に関するアンケートによる調査報告

所属機関別比率

所属	比率(%)
大学	68
公的機関	15
団体	2
企業	14
その他	1
合計	100

専門分野別比率

分野	比率(%)
ライフサイエンス	37
情報通信	9
環境・エネルギー	15
ナノテク・材料	23
ものづくり・社会インフラ	10
宇宙・海洋・その他	6
合計	100

