

オープンサイエンスの効果と課題

—新型コロナウイルスおよび COVID-19 に関する学術界の動向

池内 有為*

キーワード：研究データ公開，データ共有，プレプリント，オープンアクセス，オープンサイエンス

1. はじめに

オープンサイエンスの効果はさまざまに論じられてきた。たとえば経済協力開発機構(OECD)の報告書¹⁾では、科学研究の効率化，研究の透明性や質の向上，技術革新の加速，経済への波及効果，地球規模の課題への取り組み，共同研究の推進などが挙げられている。

折しも2019年12月以降，新型コロナウイルス(2019-nCoV)の感染が中国の武漢市から世界中に拡大して「地球規模の課題」となっている。2020年1月30日には世界保健機関(WHO)が国際的な緊急事態であると宣言した²⁾。これに対して各国・地域の研究者，助成機関，学術出版社といったステークホルダーが最新の研究成果の公開と再利用を進めて対策にあたるという，まさに「オープンサイエンス」というべき動きが起きている。そこで本稿は新型コロナウイルスに関する学術界の動向を概観した上で，今回の事例におけるオープンサイエンスの効果と課題について考察したい。なお，本稿の記述は2020年2月6日時点の情報に基づいている(情報が古い箇所もあるかもしれませんが，ご容赦を)。

2. 新型コロナウイルスに関する研究成果の公開と再利用

2.1 研究成果の迅速な共有を目指す取り組み

研究者にとって論文は研究成果を公開し，先取権を確保するためのツールである。そして評価，採用，昇進，研究資金の配分などにも関わる重要な業績である。従って，論文の出版前にその根拠となる研究データや成果を公開することは——論文の投稿時に新規性が認められない可能性や他の研究者に公開したデータを用いて論文を出版してしまう可能性があるため——難しい。しかし，これまででもジカ熱やエボラ出血熱といった感染症の国際的な流行に対処するために，論文出版前に研究データや成果を公開したとしても，賛同する出版社や雑誌で論文を出版する際には問

題としないとする取り決めが結ばれてきた³⁾。そしてWHOや公衆衛生関係者をはじめ，誰もが最新の研究成果を利用できるよう共有されてきた⁴⁾。

新型コロナウイルスに関する研究成果についても，迅速かつ幅広く共有するという声明⁵⁾が，1月31日に英国のウェルカム財団(Wellcome Trust)のウェブサイト上で公開された。現在，助成機関，学術出版社，学術リポジトリなど90以上の団体が署名している。以下では，実際に公開された研究成果や学術出版社の取り組みを紹介する。

2.2 塩基配列データ

国際的な塩基配列のデータベースであるGenBankには，1月12日から新型コロナウイルスの塩基配列データが登録されている。1月31日に米国国立医学図書館(NLM)は，「新型コロナウイルスデータのハブ」⁶⁾を公開して，GenBankに登録されたデータ一覧の提供を開始した。リストによれば，1月12日から31日までに登録されたヌクレオチドデータは30件，タンパク質データは181件であった。そして2月6日現在，ヌクレオチドデータは45件，タンパク質データは275件まで増加している。

2.3 プレプリント

生物学分野のプレプリントサーバであるbioRxivには，1月19日から31日までに31本の新型コロナウイルスに関連するプレプリントが登録された。そして2月6日現在，39本まで増加している。データの公開から，データを再利用した成果の公開に至るスピードは速い。たとえばDongらは，1月12日にGenBankに登録されたデータ⁷⁾を用いた分析結果を1月22日にプレプリントとして公開した⁸⁾。なお，後述するように既に撤回されたプレプリントも登場している。

2.4 論文と学術出版社

査読を経た論文は，プレプリントと比較すると(当然のことながら)公開までのスピードがやや落ちる。1月24日から2月6日現在までで，医学雑誌*Lancet*には5本の原著論文と3本の短報(correspondence)が⁹⁾，*New England Journal of Medicine*には3本の原著論文と2報の短報が掲載された。

2.1で触れた「新型コロナウイルス(nCoV)の発生に

*いけうち うい 文教大学文学部英米語英米文学科

〒343-8511 埼玉県越谷市南荻島 3337

E-mail: ikeuchi@koshigaya.bunkyo.ac.jp

 <https://orcid.org/0000-0002-5680-1881>

(原稿受領 2020.2.7)

Refine Results		Selected Results: 0								
		Nucleotide (78)				Protein (464)				
		Details	Accession	Release Date	Species	Length	Geo Location	Host	Isolation Source	
Virus		<input type="checkbox"/>	+	MT008022	2020-01-31	Severe acute re...	322	Italy	Homo sapiens	oronasopharynx
Nucleotide Sequence Type		<input type="checkbox"/>	+	MN933835	2020-01-24	Severe acute re...	287	China	Homo sapiens	oronasopharynx
Geographic Region		<input type="checkbox"/>	+	LR752395	2020-02-01	Severe acute re...	29872			
Host		<input type="checkbox"/>	+	MN908947	2020-01-12	Severe acute re...	29903	China	Homo sapiens	
Isolation Source		<input type="checkbox"/>	+	MN975266	2020-01-24	Severe acute re...	107	China	Homo sapiens	oronasopharynx
Collection Date		<input type="checkbox"/>	+	MT008023	2020-01-31	Severe acute re...	322	Italy	Homo sapiens	oronasopharynx
Release Date		<input type="checkbox"/>	+	MN933834	2020-01-24	Severe acute re...	29838	China	Homo sapiens	oronasopharynx
Environmental Source		<input type="checkbox"/>	+	MT020860	2020-02-05	Severe acute re...	29882	USA	Homo sapiens	oronasopharynx
Lab Host		<input type="checkbox"/>	+	MT020781	2020-02-05	Severe acute re...	29847	Finland	Homo sapiens	
Vaccine Strain		<input type="checkbox"/>	+	MN994467	2020-01-28	Severe acute re...	29882	USA	Homo sapiens	oronasopharynx

図1 NLM「新型コロナウイルスデータのハブ」⁶⁾

図2 Springer Natureの新型コロナウイルスに関する専門ページ¹²⁾

関する研究データと発見の共有⁵⁾ 声明に賛同する Wiley¹⁰⁾, Taylor & Francis¹¹⁾, Springer Nature¹²⁾ などの商業学術出版社は、それぞれ新型コロナウイルスに関する専門ページを設け、関連論文や解説記事を集約している。各出版社は、期間限定で論文をオープンアクセスにして研究や報告書への利用を認めるとしている。また、かねてより研究成果の早期共有を呼びかけていた¹³⁾ Springer Nature は、同社のプレプリントプラットフォームである In Review^{注1)} を用いた投稿を強く推奨している¹⁴⁾。

2.5 サンプル等

日本の国立感染症研究所 (NIID) は、1月31日に新型コロナウイルスの分離¹⁵⁾ に成功し、試験研究に利用する研究機関等に分与すると発表した。2月中旬から医薬基盤・健康・栄養研究所 (NIBIOHN) が細胞株の提供を開始するという¹⁶⁾。なお、NIIDの公式ウェブサイトに掲載している新型コロナウイルスの電子顕微鏡写真や細胞変性など

の画像は、商業目的や出版目的でない場合に限り使用できるとしている。

中国では、1月29日に電子商取引サイトを運営するアリババグループ (阿里巴巴集団) のクラウド部門が、新型コロナウイルスの拡散を防ぐことを目的として、各国の研究機関に AI (人工知能) 技術やツールを無償で提供すると発表した。翌1月30日には、サーチエンジンを提供するバイドゥ (百度) が、LinearFold アルゴリズムを AI 演算能力 (AI computing capacity) とともに無償提供すると発表した。これらにより、新型コロナウイルスのゲノム解析やワクチン開発などにかかる時間を短縮することが期待できるという¹⁷⁾。

3. オープンサイエンスの効果

このように、新型コロナウイルスに関する研究データやプレプリントをはじめとする研究成果の迅速な共有、すなわちオープンサイエンスの取り組みが、研究者、助成機関、学術出版社、図書館などのステークホルダーの協力によって進められている。その最大の効果は、研究開発費や人的・物的資源、時間を効率的に活用し、研究を急速に進展させていることであるといえるだろう。今後、臨床実験や新薬の開発といったフェーズも成果の共有によって加速され、感染による被害を収束させるまでの時間を短縮することができれば、オープンサイエンスの取り組みが人類の生命や健康に関わる地球規模の課題解決に貢献した好例となるだろう。

オープンサイエンスの間接的な効果として、新型コロナウイルスの感染を1日でも早く食い止めることができるならば、社会や経済への悪影響を低減することが可能になると考えられる。また、新薬をはじめとする対策によって経済的な効果も生じるだろう。研究データや成果を商用利用することに対しては抵抗もみられるが、ゲノム研究に関するカナダ、米国、および英国のガイドラインを分析した Caulfield らは、研究の商業化とオープンサイエンス政策は必ずしも折り合わないとはいえないと指摘している¹⁸⁾。

2017年6月にSpringer Natureのデータジャーナル *Scientific Data* は、同誌に投稿されたデータセットは機密データを除いて商用利用の制限を認めないとする(=商用利用OKとしなければならないという)方針を公開した。共有されたデータが社会や経済にとって利益をもたらした例として、ヒトゲノムプロジェクトや米国のLandsat衛星のデータは、推定数十億ドルを生み出したことを挙げている¹⁹⁾。

4. オープンサイエンスの課題

本件は、データ公開やフォーマットの標準化が慣習となっている分野であるからこそ、オープンサイエンスの効果が発揮されたといえる。一方で、課題も散見された。ここでは結果の誤用とアクセスツールの整備について述べている。

4.1 研究成果の誤用・悪用

Pradhanらが1月31日にbioRxivに登録した“新型コロナウイルスにHIVウイルスに類似したタンパク質が含まれている”とするプレプリント²⁰⁾は、研究者から内容に対する複数の批判的なコメントを受けた。さらに、フェイクニュースの根拠として使われ、拡散される事態に発展した。結果、Pradhanらは公開からわずか2日後の2月2日にプレプリントを撤回した。

前半部分、すなわち研究の内容に対する批判が公開で自由に行われることは、研究の透明性や信頼性の担保となるためオープンサイエンスの効果として評価できる。一方、研究成果の誤用については国内外の調査において研究者によるデータ公開の障壁であることが示されている²¹⁾²²⁾。本件は意図的な誤用であると考えられるため、一層悪質であると言えよう。Pradhanらはスピーディな撤回により問題を最小限に抑えたといえるかもしれないが、研究成果の悪用は研究者のみならず社会にとっても深刻な問題であるため、対策を講じる必要があるだろう。WHOはパンデミックならぬ“infodemic(情報伝染病)”を引き起こさないよう対策を講じることを²³⁾、国際連合(UN)は根拠なき差別の終結を求めるとの見解を²⁴⁾、それぞれ発表した。

4.2 研究成果へのアクセスツール

2.2で紹介したように、NLMは新型コロナウイルスに関する塩基配列データを整理したリストを公開した。また、文献情報をDisaster Lit^{注2)}やPubMedからまとめて検索できるようにリンクを提供している。1月14日には、医学件名標目表(MeSH)の補足概念用語(Supplementary Concept Records, SCRs)に“Wuhan coronavirus”を追加しており²⁵⁾、適合度の高い文献の検索を可能にした。

新型コロナウイルスに関する研究データは2019年6月号で紹介したGoogle Dataset Search^{注3)}、DataCite Search, Elsevier DataSearch(ベータ版)でも検索できる。ただし、これらは分野を横断した広範なデータを対象としていることや収録対象が異なることから、“Novel

Coronavirus”, “Wuhan Coronavirus”, “2019-nCoV”など同一のキーワードを用いて検索しても結果はバラバラで、本稿に用いたデータはうまくみつけられなかった(※筆者の知識・能力不足も影響していると思う)。今後、データのアクセスツールがどのように発達・融合していくのかは未知数であるが、ツール間の役割分担や各ツールを適切に使いこなすための知見の整理が課題であると感じた。

5. おわりに

オープンサイエンスの動向について、佐藤翔さん(同志社大学)、林豊さん(九州大学(当時)・現国立情報学研究所)、尾城孝一さん(国立情報学研究所)とともに自由に書かせていただいた本リレー連載。残念ながら、今回で一旦休止となる。筆者は研究データにまつわる基本的なトピックを取り上げてきたため、他の連載陣と比較して速報性に欠けるのではないかと心配であったが、最後に世界規模で起きている「オープンサイエンスのいま」について書くことができた。おそらく本稿の校正時点、公開時点でも新たな展開が起きていることだろう^{注4)}。

興奮しすぎていつにも増して遅れた原稿を辛抱強く待つて下さった編集委員長であり企画主でもある南山泰之さん(国立情報学研究所)と編集事務局のTさんに心からの感謝を申し上げて筆を置きたい。もちろん読者の皆様にも。2年間ありがとうございました。

註・参考文献

- 注1) BMCなど、Springer Natureの雑誌に投稿するプレプリントのためのプラットフォーム。投稿、査読、編集プロセスを公開している。
<https://www.researchsquare.com/publishers/in-review>, (accessed 2020-02-06).
- 注2) 災害時に医療や公衆衛生に関するガイドライン、報告書、図書、ウェブページなどの情報を提供するデータベース。NLMは、検索式“(coronavirus AND novel) OR Wuhan”による結果のリンクを提供している。
[https://disasterinfo.nlm.nih.gov/search/?q=\(coronavirus%20AND%20novel\)%20OR%20Wuhan](https://disasterinfo.nlm.nih.gov/search/?q=(coronavirus%20AND%20novel)%20OR%20Wuhan), (accessed 2020-02-06).
- 注3) 2019年6月号の時点ではベータ版であったが、2020年1月23日に正式公開された。
<https://datasetsearch.research.google.com>, (accessed 2020-02-06).
- 注4) 2月11日に、WHOが新型コロナウイルスによる感染症の名称を「COVID-19」とすると発表したため、本稿のタイトルに追記した。
<https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>, (accessed 2020-02-13).
- 1) OECD. Making Open Science a Reality. OECD Publishing. 2015, 108p, (OECD Science, Technology and Industry Policy Papers, no. 25).
<https://doi.org/10.1787/5jrs2f963zs1-en>, (accessed 2020-02-06).
- 2) 2019-nCoV outbreak is an emergency of international concern. World Health Organization, 2020-01-31.
<http://www.euro.who.int/en/health-topics/emergencies/pages/news/news/2020/01/2019-ncov-outbreak-is-an-emergency-of-international-concern>, (accessed 2020-02-06).
- 3) 公衆衛生の危機に際して、データ共有を進める声明(仮訳)。

- 国立研究開発法人日本医療研究開発機構, 2016-02-16. <https://www.amed.go.jp/news/topics/20160216.html>, (参照 2020-02-06).
- 4) 日本医療研究開発機構 (AMED) は新型コロナウイルスの流行に対処するため, 新型コロナウイルスに関連する研究成果とデータを広く迅速に共有する声明 (令和 2 年 1 月 31 日) に署名しました. 国立研究開発法人日本医療研究開発機構, 2020-02-03. <https://www.amed.go.jp/news/topics/20200203.html>, (参照 2020-02-06).
 - 5) Sharing research data and findings relevant to the novel coronavirus (nCoV) outbreak. Wellcome Trust, 2020-01-31. <https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak>, (accessed 2020-02-06).
 - 6) Novel coronavirus data hub: Search, retrieve, and analyze 2019-nCoV Genbank data. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Wuhan%20seafood%20market%20pneumonia%20virus,%20taxid:2697049, (accessed 2020-02-06).
 - 7) Wu, F., et al. Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome. GenBank, MN908947.3. <https://www.ncbi.nlm.nih.gov/nucleotide/MN908947>, (accessed 2020-02-06).
 - 8) Dong, N., et al. Genomic and protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China. *bioRxiv*, 2020-01-22. <https://doi.org/10.1101/2020.01.20.913368>, (accessed 2020-02-06).
 - 9) 2019-nCoV Resource Centre. *The Lancet*. <https://www.thelancet.com/coronavirus>, (accessed 2020-02-06).
 - 10) Novel Coronavirus Content Free to Access. Wiley Online Library. <https://novel-coronavirus.onlinelibrary.wiley.com>, (accessed 2020-02-06).
 - 11) Coronavirus reading list. Taylor & Francis Group News, 2020-01-27. <https://newsroom.taylorandfrancisgroup.com/coronavirus-reading-list-for-journalists/>, (accessed 2020-02-06).
 - 12) Novel Coronavirus (2019-nCoV). Springer Nature. <https://www.springernature.com/gp/researchers/campaigns/coronavirus>, (accessed 2020-02-06).
 - 13) Mitchell, A.; Swaminathan, S. Maximising the benefits of early sharing. Springer Nature, 2019-05-23. <https://www.springernature.com/gp/advancing-discovery/blog/blogposts/maximising-the-benefits-of-early-sharing/16744616>, (accessed 2020-02-06).
 - 14) Press Releases: Novel Coronavirus - (2019-nCoV). Springer Nature, 2020-01-30. <https://group.springernature.com/gp/group/media/press-releases/springer-nature-coronavirus/17607604>, (accessed 2020-02-06).
 - 15) 新型コロナウイルス: 国立感染症研究所が開発した細胞で分離に成功. 国立感染症研究所, 2020-01-31. <https://www.niid.go.jp/niid/ja/basic-science/virology/9369-2020-virology-s1.html>, (参照 2020-02-06).
 - 16) 新型コロナウイルスの分離に利用可能な細胞株の提供開始について. 国立研究開発法人医薬基盤・健康・栄養研究所, 2020-02-04. <https://www.nibiohn.go.jp/information/nibio/2020/02/006237.html>, (参照 2020-02-06).
 - 17) Alibaba and Baidu offer AI gene sequencing tools to help fight the China coronavirus outbreak. *KrASIA*, 2020-01-31. <https://kr-asia.com/alibaba-and-baidu-offer-ai-gene-sequencing-tools-to-help-fight-the-china-coronavirus-outbreak>, (accessed 2020-02-06).
 - 18) Caulfield, Timothy; Harmon, Shawn H. E.; Joly, Yann. Open science versus commercialization: A modern research conflict? *Genome Medicine*, 2012, vol.4, no.2, 17. <https://doi.org/10.1186/gm316>, (accessed 2020-02-06).
 - 19) Editorial: Open for business. *Scientific Data*. 2017-6-13. <https://doi.org/10.1038/sdata.2017.58>, (accessed 2020-02-06).
 - 20) Pradhan, P., et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*, [retracted 2020-02-02]. <https://doi.org/10.1101/2020.01.30.927871>, (accessed 2020-02-06).
 - 21) 池内有為, 林和弘, 赤池伸一. 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, 108p. <https://doi.org/10.15108/rm268>, (参照 2020-02-06).
 - 22) 近藤康久. レポート紹介「地球環境研究におけるオープンデータ: バルモント・フォーラムによるオープンデータ調査」. *情報管理*, 2016, vol.59, no.4, p.250-258. <https://doi.org/10.1241/johokanri.59.250>, (参照 2020-02-06).
 - 23) Novel Coronavirus(2019-nCoV) Situation Report – 13. World Health Organization. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf>, (accessed 2020-02-06).
 - 24) Coronavirus: UN health agency moves fast to tackle 'infodemic'; Guterres warns against stigmatization. *UN News*, 2020-02-04. <https://news.un.org/en/story/2020/02/1056672>, (accessed 2020-02-06).
 - 25) "Wuhan coronavirus", MeSH Supplementary Concept Data 2020. U.S. National Library of Medicine. <https://meshb.nlm.nih.gov/record/ui?ui=C000656484>, (accessed 2020-02-06).

Series: Current trend of open science: Effects and challenges of Open Science — Academic movements related to the Novel coronavirus (2019-nCoV) and COVID-19. Ui IKEUCHI (Bunkyo University, Faculty of Language and Literature, Department of English Language and Literature, 3337 Minami-Ogishima, Koshigaya, Saitama 343-8511)

Keywords: Open Research Data / Data Sharing / Preprint / Open Access / Open Science