

ほらいずん

# 研究データのオープンアクセスを担保する 機関リポジトリの展開

データ解析政策研究室 客員研究官 引原 隆士

## 【概要】

本稿は、論文が根拠とする研究データについて、オープンアクセスの考え方を整理する中で、研究者が所属する機関の機関リポジトリの在り方について整理したものである。特に、機関リポジトリが研究データのオープン化とその運用の鍵を握っていることへの理解と、それにより研究者自身が行動変容を起こす必要があること、そしてそれを受けて機関リポジトリが相互にリンクしていく仕組みの構築が重要であることを述べている。研究データを所有し公開することの権利、またそのデータの重要性を導き出した論考、コード等の重要性を公知とすることは、その研究領域を広げると同時に研究教育に資するところが大である。それらの取組の事例として、京都大学で進めている研究DXを支えるデータ運用支援基盤センター設置とその取組について示し、最後に、研究データのオープンアクセスが目指すべき方向性について触れる。

キーワード：オープンサイエンス，オープンアクセス，オープンデータ，機関リポジトリ

## 1. はじめに

公的な研究助成機関 (FA) による競争的研究費は、我が国において研究者の基礎研究から応用研究、研究の芽生えから発展までを広く支援する研究助成の根幹をなす。大学や国研における基盤的な運営資金の削減を受けて、人件費や施設整備等の固定費が大きい各機関では研究費を確保することが困難となっていることから、競争的研究費の獲得が研究の推進には避けて通れない状況となっている。統合イノベーション戦略推進会議 (2024年2月16日) において、「公的資金のうち2025年度から新たに公募を行う即時オープンアクセスの対象となる競争的研究費を受給する者 (法人を含む) に対し、該当する競争的研究費による学術論文及び根拠データの学術雑誌への掲載後、即時に機関リポジトリ等の情報基盤への掲載を義務づける。」ことが決定された<sup>1)</sup>。今後それを受けた施策が実施されていく中で、研究者が所属する機関においては、成果論文とその根拠データの公開を支援する体制及び基盤の確立と、そのバックデータのアーカイブ及び研究資源としての活用を確立していく対応

が喫緊の課題となっている。

既に、論文の公開は preprint による公開、APC を支払って OA 化するハイブリッドジャーナル、OA ジャーナルによる Gold OA、そして機関リポジトリ上による著者最終稿の公開 (多くはエンバーゴ期間が設定される) と、論文の著作権と出版権などの関係に基づく多様な選択肢があり得る。これら個々の成立の過程をここで述べることはないが、この過程は研究者と出版社のインターネット上における論文公開の主権を巡る争いでもある。当事者でなければそれらのどの方法が取られても論文がオープン化されていることに変わりはなく、学術情報の流通に寄与している。しかし、それぞれの意味は、論文の成立日時一つ取っても大きく異なる。この議論において欧州における Plan S の提案<sup>2)</sup>、米国 OSTP による即時 OA の指示<sup>3)</sup>等は、公的資金を受けた研究成果はその原資となる税金の納税者たる公衆に還元すべきものであるという基本理念に基づいており、商業出版の営利活動であっても研究活動を支えている研究資金の原資の属性に基づいて従うべきものとの考えによる。出版社は営利活動がよって立つ学術の義務においてその要求

に従うべきという重要な視点を伴っている。

本稿は、以上の背景を踏まえ、論文が根拠とする研究データについて、オープンアクセスの考え方を整理する中で、研究者が所属する機関の機関リポジトリの在り方について整理する。機関リポジトリの研究データのオープン化における重要性と同時に、研究者自身が行動変容を起こす必要があること、及び機関リポジトリが相互にリンクしていく仕組みの構築が重要であることを述べる。取組の事例として、京都大学が進めている研究 DX を支えるデータ運用支援基盤センター設置とその取組について示す。最後に、研究データのオープンアクセスが目指すべき方向性について触れる。

## 2. 研究論文と研究データ

今回の議論の根底をなす重要な要素として、そもそも論文とは何かから始める。分野ごとの差異はあるにしても、自然科学においては公開されたデータを基に理論、数値計算、実験により検証し、少なくともそのいずれか二つの組合せによって理（ことわり）の蓋然性を導くものである。未だ定義されていない現象に関しては、対象とするパラメータ領域において現象等の発生の有無を独立かつ網羅的に確認し、検出された現象の再現性を保証した上で、さらにその現象が他の同領域で生じる現象と異なることを特定することが求められる。従って、研究データはその論文において初めて公開される場合が多い。一方、人文社会学においては、誰もが利用でき公開されている検証可能な資料・データに基づき、新たに見いだされた解釈を論じる場合がある。例えば、それらの資料・データに対して他の解釈の可能性を排除した上で、選択が論じられる必要がある。これらの形式は一部の例にすぎないが、少なくとも論文は誰もが確認できる公開済みの資料・データを根拠として、人類がこれまで知らなかった普遍的な理解にたどり着き、知を広げたことを検証できる形で示す論述であらねばならない。もちろんその資料・データの正当性が保証されていることが前提であり、創作による物語ではない。

さて、記述された論考の論文としての要件は、独創性、再現性、妥当性である。工学ではさらに有用性などが付け加えられる場合もある。従って、必ずしも論文の要件は絶対的ではない。それは論文誌の

編集ポリシーによる。指定された要件が満たされるかどうかは査読システムで検証される。あえて言えばステークホルダーとなる研究者全ての性善説に基づいて成立するシステムである。そしてそれが論文誌がよって立つ前提である。これが全てボランティアと少数のコミュニティ内の相互研修の場であれば一般に実害はない<sup>注1</sup>。コミュニティ内の検証を経てデータが公開され、取得手法の真偽が検証されたものであれば、そのデータに基づく論文に対しては議論の余地なく再現性は保証されると考えられる。従って、次の段階に研究を深化させることができる。多くの公的資金を受けた研究成果はこの過程を振興するものでなければならず、その価値観が共有されていなければならない。成果として必ずしも既存の結論が正しいことを保証する必要はないが、新たに論考のよりどころとするデータが、その出所と取得手法を明示した上で公開されなければならない。

論文誌は、査読というプロセスを経ながらも実際には玉石混交で論文を通し、その示された条件において論考の成立要件が後から否定された場合には、否定された論考の試みとしての位置づけのみを残してとう汰されるとするのがこれまでに取られてきた方法である。

繰り返しになるが、近年学術論文自体が公的資金を受けた場合には公的な研究の成果と考えられてきた。論文とはその主張のオリジナリティ、再現性、妥当性を担保する根拠データが全てその中に記述され、誰もがそのデータにアクセスして再現性を確認でき、他の主張との差異を確認できるものであるとするのが共通理解である。論文と異なりデータの扱いに関して明確な議論がなされてきたとは言えない。その一因は、研究者が研究を独占的に継続できる権利として根拠となる研究データ所有の権利が、著作権や特許のような制度で守られたものになっていないことにより、研究資源としての専有、もしくは関係者外非公開が許されてきた過去の経緯がある。しかしながら根拠データが非公開である場合、その真偽を確認する手段がないばかりか、客観性を担保できないため研究公正の意味においても確認ができない。

研究データが、何時、どこで、誰によって、どのような条件下で、また方法で取得されたものであるかは、そのデータの来歴を示し再現性を担保するデータ固有の情報により担保される。これらをメタデータと

注1 しかし問題の一つは、「研究業務の中で時間を割いて査読をする研究者は、これらのシステムが出版社、学会の事業の収益源となっているにもかかわらず、無償の奉仕者と扱われ、収奪の対象になっているという認識がない。」ということである。

呼ぶ。しかしこれは論文等のカテゴリーを類別する書誌的メタデータとは大きく異なる。研究のプロセスの中で、物理的取得条件、変換コード、キュレーション、さらには統計処理の方法などが含まれ、その上で論文の論考の根拠として必要十分な情報が含まれていなければならない。それを結果から見た書誌的メタデータと同様に扱うことは危険である。すなわち、同じデータから複数の論考が生まれていくということを見ると、方向を逆にした来歴が重要となる（図表1）。これは、書籍、論文誌が物理的な形態と著作権で守られる記述の固有性、独自性を有することは異なり、データ+メタデータとしてそのオリジナリティや普遍性を確保しなければならないことを意味する。

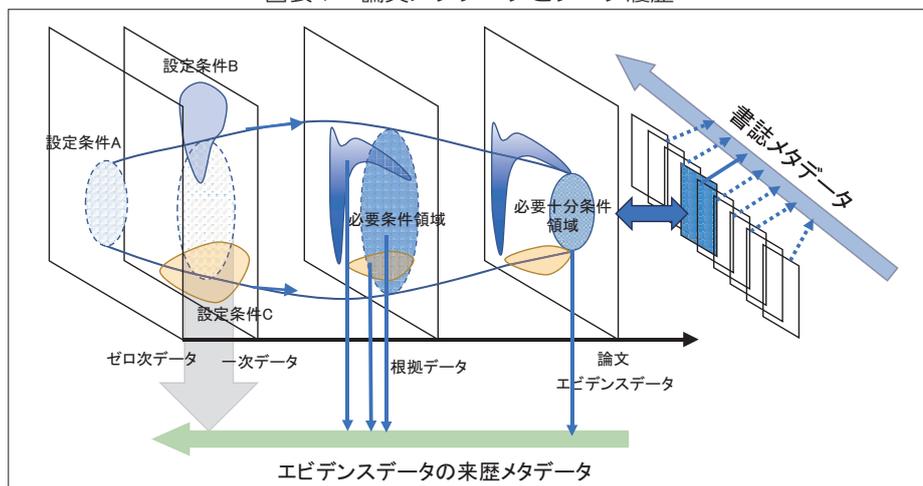
公開（オープン）、非公開（クローズ）、限定公開（シェア）という研究データの扱いに関しては、機関における研究推進の観点に立脚し、研究機関が一方的に研究者に指示するだけではなく、研究者に併走して研究のサイクルを回し、機関リポジトリ等を介して研究者と研究コミュニティにおける研究活動を活性化するために、ステークホルダーである研究者自身が考えるべきことである。商業出版や企業活動の自らの利を正当化する論理にくみすることは避けるべきことだということは明らかである。この点が忘れられると、義務の実施の確認や無用な採点主義の商品導入を生み、本末転倒な義務化への対応・あるいは回避手段を生むことにもなる。英語を中心とする成果論文の出版において我が国が後手に回った経緯から見ても、商業的に資源生産を担うステークホルダーの連携を阻み、競争させる仕組みを作り上げた現状を、冷静に判断する必要がある。

### 3. 研究のライフサイクルと研究成果のオープンアクセス

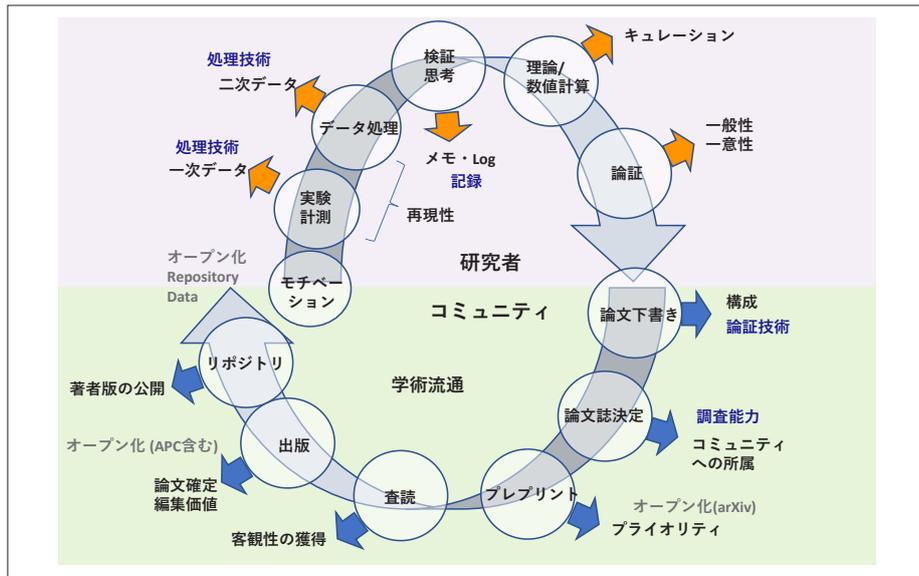
研究のライフサイクル（図表2）は、自立した研究者が定常的に研究を推進する際の作業手順に基づく。しかしながら、研究者が最初からこのサイクルを理解して進めているわけではなく、博士学位などの取得の過程で指導者から教育を受け、経験的に習得している。妥当な結果が得られるまでの繰り返しの実験や計算、それに対する指導者の判断を伴うキュレーション、そしてその後の論文執筆と研究成果の公開手順などである。STEM分野においては、多くの大学で博士学位の取得が少なくとも複数編の学術原著論文の出版を経ることが条件として運用されているのは、このサイクルの習得が学位指導の成果であろうとの認識があることによる。学位認定は各大学、大学院の専権事項であり、標準化されたものはない。そのため、コミュニティの評価判断しやすい論文数を条件とするという判断が客観的指標として使われている。文部科学省が定めた学位規則において、博士学位論文は本紙もしくは要旨についてエンバーゴ期間を認めた上でインターネット公開を義務化したのは2013年4月のことである。しかしながら、学位論文のインターネット公開義務と論文の著作権との関係において、慣行的に行われてきた商業出版と義務化によるオープンアクセスの間で判断の揺れを生じている。ここに論文とデータが関わる問題の縮図がある。

従来、学位に関する取扱いは大学における教務、準貴重図書管理としての図書館の業務とされ、研究に関わる処理は研究者の裁量とされてきた。このことが図書館が機関リポジトリにより学位論文を管理・運用している理由である。学位論文電子公開に当たって、これを教務あるいは図書館が自らの業務ではない

図表1 論文メタデータとデータ履歴



図表2 研究のライフサイクル



日本学術会議オープンサイエンスの深化と推進に関する検討委員会、  
提言「オープンサイエンスの深化と推進に向けて」（2020年5月28日）付録図表3を著者許諾の上改訂

と判断した大学もある。その結果、著作権問題や特許申請を理由とする学位論文公開回避の抜け道を学生、指導教員に許し、不明確な扱いを残してきた。出版社の著書としての販売を理由に非公開化する例も多々見られる。これらの公的な学位成果が個人及び出版社の権限に移譲されることが許されているか否かに、誰も正確に答えていない。しかもどの学位論文でも著者の承認なく50%までは複製できることは知られている。従って、根拠データはテキストの形でシェアされている。その結果、投稿論文と学位論文でデータの使いまわしという訴えがなされることもある。そもそも、学位論文の研究成果とは公開論文なのであるか。公開論文はその体系的な研究の一部にすぎず、それを形式的に束ねて学位とするという手法ではなく、その結果として一連のデータの関係に基づいて、新しい知見を創出したかどうかである。従って、特許のために非公開とするにしても、その手続は学位取得プロセス前に行うべきで、公開原則の学位においてそれを理由とすることも、明確な判断が放置されている。

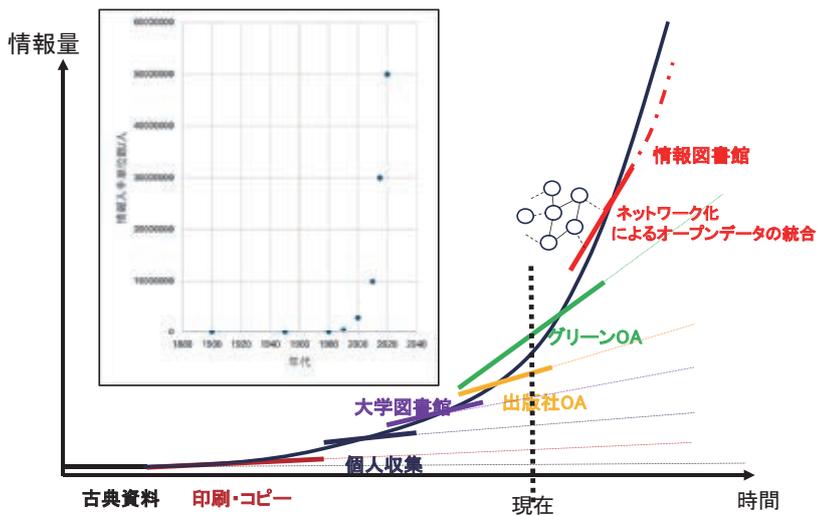
研究成果のオープンアクセスについても、既に出版されて公開されている論文をなぜオープンアクセスにするのかという点に関して、同様に現時点で明確な説明ができていない。今回の政府の決定を受けた研究者への要求が、この点を、不要な作業要求と捉えられる素地となることは否めない。すなわち、研究の推進とは公的資金を受けた研究を各機関が系統的に扱うもので、その（デジタル時代の）研究サイクルを完成することを求めたということが認識されていない。多くの大学では研究サイクルの支援及び事務処理

が別の部署や研究科等で連携なく個別になされ、それに慣れた研究者がその求めの必要性を自らの業務として不要なものとする危険性がある。研究公正などの確認は研究のサイクルの中で行われれば苦もない話であるが、別の部署がその業務として研究の流れと関係なくエビデンスを求めてくることは、研究者からしてみれば煩わしいだけである。それと同様に、流れを知らずに公開を義務化する要請を大学が研究者に出すことは同様に想像できる。これは、大学という本来研究者を守るべき組織の中で生じる業務の局所最適化の弊害で、本来の研究組織の在り方も含めて設計する必要がある。

#### 4. オープンアクセスが生み出す環境とその意義

研究者が自ら検索し、研究活動に資することができる学術情報について考察する。容易に想像できるように、インターネットが普及する前後を知る世代は、手作業と紙媒体の移動を伴う古典的通信手段に頼っていたが、図表3に示すように、概数的な数値を用いても、その研究資料数の増加は明らかな指数的爆発を示している。その中で、研究者、所属機関等がコストと対価のバランスを求めることは不可能である。図書館は、書庫（ストレージ）機能と世界への窓（インターフェース、ルータ）の機能として各機関で役割を担ってきた。しかし現在、それらは巨大な学術情報ネットワークの中の一つの点に過ぎない。従って、それら機関の図書館の資料数という個数管理で

図表3 学術情報の指数的爆発



作図に当たって概算値は以下の通りとした。個人収集の平均 <100、国会図書館 1,000 万、京大図書館機構 700 万、部局図書館 数万～50 万、日本の購入論文誌タイトル 1 万、OASPA @ 2022 110 万論文、米 国国会図書館資料 2 億弱、世界の蔵書数～ 数十億。  
1900 年 100 (個人)、1950 年 1,000 (個人)、1980 年 10,000 (図書室)、1990 年 500,000 (複写)、2000 年 3,000,000 (論文数 @ 図書館)、2010 年 10,000,000 (コンテンツ数 @ オンライン)、2015 年 30,000,000 (ネットワーク)、2020 年 50,000,000 (OA)。数値の単位は資料点数。

きるものには限界がある。既に多くの研究資源は図書館の中になく、巨大なネットワークのどこかの記憶媒体の中に見いだせる情報が前提となる。その中で、ノードにとどまるデータは流通がなく、これまでに知られている結果としては不十分となる。すなわち論考の立脚点となる根拠データがオープンアクセス可能で、流通可能になっていることが求められる。

上述したオープンアクセスの環境をどのように構築するかという点で、研究大学はそのための対応を取ることに未だ積極的でない。それは大学自身が研究者の運営する組織で、その運営が自らの過去の経験に基づく保守的なステークホルダーとなる場合が多いからである。図表3のように研究者が扱う学術資料数の指数的増加に対して、数年前の経験ですら意味を持たなくなる。加速していく時代にあって、過去を問うのではなく、今から後を問うということが重要となる。そのために必要なことは、これから後の学術に対しては、より公開されたデータに基づく判断を要件とすることを課すと同時に、過去のものへは他者による当時とは異なる環境による再検討結果を踏まえて、研究のスパイラルな展開を求めることが正しい。この合意なしに、義務的なオープンアクセスの要請は成立しない。

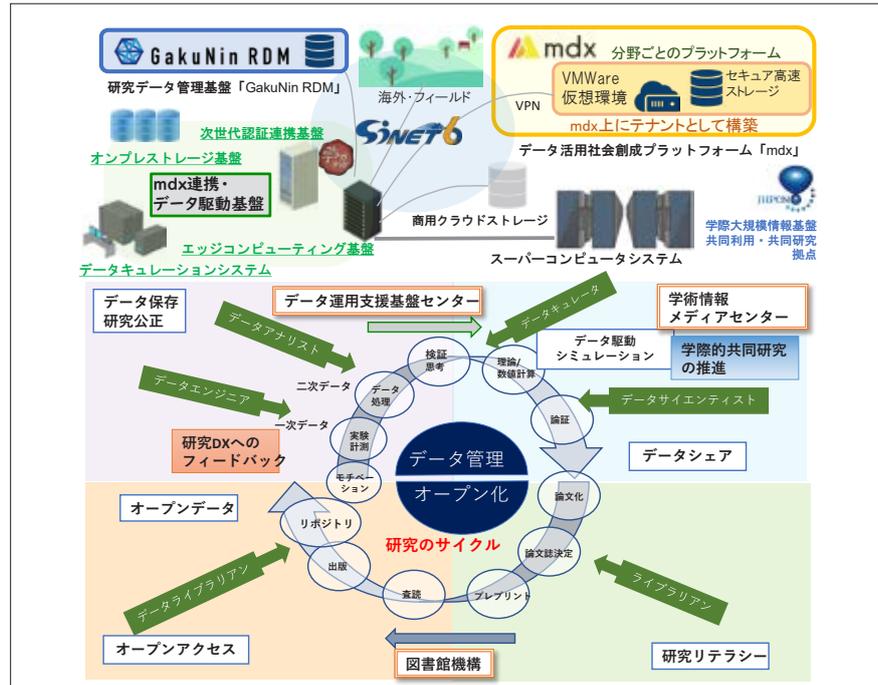
## 5. 京都大学における取組事例と研究データのオープンアクセスが目指すべき方向性

本稿の最後に京都大学の事例を示す。図表2の研究のライフサイクルを理解した上で、オープンアクセスを個々の研究活動のステップとして置くことができるよう、学内の研究者の多様な意見を聞き取った上で進めてきた、「データ運用支援基盤センター」の創設

である。これは従来の全学機能組織である情報環境機構の改組に基づく(図表4参照)。

京都大学の機関リポジトリ KURENAI は、資料数 159,000ITEM 数(本稿執筆時点)が Google Scholar でオリジナル資料としてカウントされている機関リポジトリである。2024年3月時点で、世界第6位の機関リポジトリとなっている。これ以外にグリーンOAの論文著者版が公開されている。長期の戦略を立てて進めてきたブランディングの成果であり、研究ライフサイクルの最後のオープン化部分を担っている。同じく、学内所蔵の資料をデジタル化し、世界標準フォーマット(IIIF)によるデジタルアーカイブとしてワンストップで公開している。その中には国宝などの貴重資料も含まれ、研究の創始を促すためのエビデンスとなるオープンアクセス資料となる。一見それらは別物のように考えられてきたが、データ運用支援基盤センターの活動ではストレージと計算資源の提供を含めて体系的に管理・運営する。学内向けだけでなく学外との連携を想定した次世代の認証系も構築する。そのため、研究のライフサイクルを支える情報基盤を連携し、情報環境機構、学術情報メディアセンター、図書館機構を、サイバーだけでなくフィジカルにもつなぐ。また、支援組織としてのコンサルタントチーム(情報系スタッフ、図書系スタッフ、URA、データエンジニア等)が研究者に伴走し、個々のフェーズは専門職がサポートする。研究データは研究現場にある。研究者が望むのはそれを尊重して、命じられるのではなく、より良い形への手引をしてもらえることである。これらを同時並行して運用する中で、データの生成と共有、そして提供する環境で研究の融合を可能にする発見性を与えるプラットフォームを形成することが機関リポジトリの次の展

図表4 京都大学の取組の事例



開と考えている。

ここに示した事例は、学術論文等のオープンアクセス化をきっかけにして、研究者が研究の主導権を取り戻すために所属機関が取り組むべき体制の手がかりを示したものである。すなわち、研究を実施する研究者だけでなく支援者を含む関係者が、研究のライフサイクルを理解し、研究の開始から成果の公開までのプロセスに伴走する形の運営を追求することである。明

らかに今ではもう一人で全ての研究プロセスを推進することは難しい。ともすると研究費の取得、繰り返し返される報告要請、さらには研究公正に伴う煩雑な事務作業が丸投げされ、研究すること自体が罪悪であるような扱いが横行している。この根本問題は研究プロセスを把握せず研究者に複数の部署がその場その場で行う業務のシステム化の欠如である。この解消こそが本来の研究成果のオープン化の目指す方向である。

参考文献・資料

- 1) 統合イノベーション戦略推進会議, 学術論文等の即時オープンアクセスの実現に向けた基本方針, [https://www8.cao.go.jp/cstp/oa\\_240216.pdf](https://www8.cao.go.jp/cstp/oa_240216.pdf)
- 2) European Science Foundation, Plan S – Principles and Implementations <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>
- 3) カレントアウェアネス, 「米国大統領府科学技術政策局 (OSTP)、連邦政府が助成した研究のオープンアクセス出版のための資金調達メカニズムに関する報告書を公開」 (2023.12.18). <https://current.ndl.go.jp/car/202480>