

ほらいずん

# 研究データの公開と論文のオープンアクセスに関する実態調査 ーオープンサイエンスの課題と展望ー

科学技術予測センター 客員研究官・筑波大学大学院図書館情報メディア研究科 池内 有為  
科学技術予測センター 上席研究官 林 和弘

## 【概要】

公的資金による研究の成果である論文やデータを公開し、学術関係者のみならず、企業や市民による利活用を可能にするオープンサイエンスの政策が世界的に推進されている。オープンサイエンスの実現によって、科学の発展やイノベーションの創出が期待されている一方で、データの公開については分野ごとの特性をふまえる必要があると指摘されている。

そこで科学技術・学術政策研究所（NISTEP）科学技術予測センターは、データ公開を中心とした日本のオープンサイエンスの実態と課題を明らかにするため、2016年11月から12月にかけて科学技術専門家ネットワークを活用したアンケートシステムによる調査を実施した。その結果、回答者1,398名のうち51.0%がデータの、70.9%が論文の公開経験を有していた。データを公開しようとする場合のリソースとして、79.2%が人材、74.5%が資金、74.1%が時間について「不足/やや不足している」と認識しており、公開の懸念として、87.8%が引用されずに利用される可能性を、84.6%が先に論文を出版される可能性を「問題/やや問題である」と認識していた。回答者の75.8%は公開されているデータを入手した経験を有していたが、利用料金や利用者登録などが問題であると考えていることが明らかになった。

## 1. はじめに

オープンサイエンスとは、“幅広い分野の公的資金による研究成果（論文や関連するデータセット等）に学術関係者だけでなく、民間企業や一般市民が、広く利用・アクセスできるようにする”<sup>1)</sup>取組である。論文のオープンアクセス（以下、「OA」）や研究データの公開と利活用によって、新たな科学の発展やイノベーションの創出、研究の透明性の向上などが期待されることから、G7科学技術大臣会合をはじめとする国際組織や各国の政府機関がオープンサイエンスの推進を表明している。第5期科学技術基本計画（2016年度～2020年度）<sup>2)</sup>では、“国は、資金配分機関、大学等の研究機関、研究者等の関係者と連携し、オープンサイエンスの推進体制を構築”するとともに、“公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢とする”と述べられている。

研究データの公開（以下、「データ公開」）は、2013

年にG8科学技術大臣会合の共同声明<sup>3)</sup>で言及されたことを契機として議論が加速した<sup>4)</sup>。2017年4月には科学技術振興機構が『オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針』<sup>5)</sup>を公開し、助成金を獲得したプロジェクトに対して研究開始までにデータマネジメントプラン（DMP）の提出を求めるとともに、データ公開を推奨している。また、学術出版社も論文のエビデンスとなるデータ公開を求めることが増えつつあり、例えばSpringer Nature社は2016年12月に<sup>6)</sup>、Wiley社は2017年9月に<sup>7)</sup>、それぞれが出版する雑誌に適用する新たな「データ共有ポリシー」を発表した。データ公開をデータ出版（data publishing）<sup>8)</sup>と呼び、論文の出版と同様に扱ったり、引用したり、業績や評価の対象としようとする動きもみられる。論文や引用情報のデータベースであるWeb of Science<sup>9)</sup>は2012年からData Citation Indexを公開してデータ引用の追跡を可能にしており、Scopus<sup>10)</sup>は論文の根拠となる研究データの検索ツールを強化している<sup>11)</sup>。

しかし、データの公開については分野によってデータの種類や機密性、取扱いの慣習などが異なるため、それぞれの特性をふまえた政策が必要であると指摘されている。2015年に公開された内閣府による報告書では、“各省庁等のステークホルダーは、オープンサイエンスを推進すべき領域、プロジェクトを選定し、研究活動上の利益・損失や研究途上の取扱い及び機微の判断など各分野の専門家・研究者、技術者の意見を十分に取り入れ、その分野の活動・研究成果が最大化されることを旨として、オープンサイエンス実施方針を定める”<sup>12)</sup>と述べられている。

そこで科学技術・学術政策研究所（NISTEP）科学技術予測センターは、政策立案や研究マネジメントに資することを目的として、日本の研究者による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に、アンケートシステムによる調査を実施した。調査期間は2016年11月30日から12月14日であり、調査対象は、大学、企業、公的機関・団体に所属する研究者や専門家、技術者などによって構成される約2,000名の科学技術専門家ネットワークである。

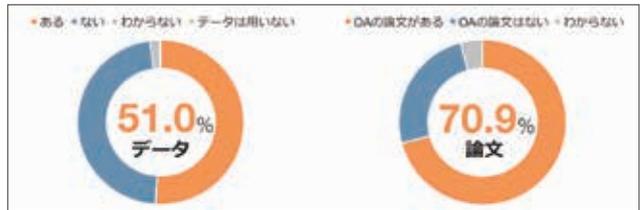
本稿では、1,398名（回答率70.5%）の有効回答を分析した結果から、(1) データと論文の公開状況と阻害要因、(2) データ公開の支援に関する検討、(3) 公開データの活用と課題について報告する。調査の詳細や全ての結果については、調査資料として公開されている報告書<sup>13)</sup>を御参照いただきたい。

## 2. データと論文の公開状況と阻害要因

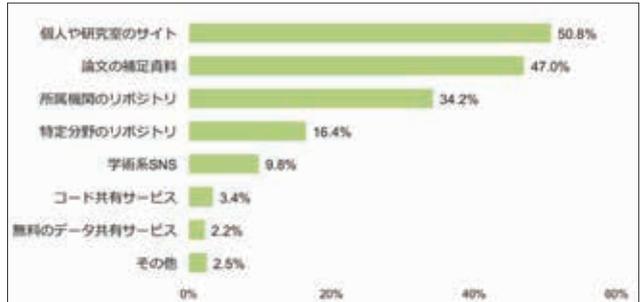
データ（研究のために収集・作成・観測したデジタルデータで、論文など研究成果の根拠となるもの）の公開経験がある回答者は713名（全体の51.0%）、OA論文がある回答者は991名（70.9%）であった（図表1）。データとOA論文両方の公開経験がある回答者は568名（40.6%）、いずれもない回答者は214名（15.3%）である。なお、多分野の研究者を対象とした先行調査によるデータ公開率を確認すると、Tenopirら（2010）<sup>14)</sup>による調査では36%（回答者数1,329名）、Wiley社による調査（2014）<sup>15)</sup>では52%（同2,250名以上）、うち日本の研究者は44%（日本の回答者数不明）、Berghmansらによる調査（2017）<sup>16)</sup>では66%（回答者数1,162名）であった。

データ公開について尋ねる際には、回答者の認識によるずれを防ぐため図表2の7種類の公開方法を複数選択方式で示し、これらの選択肢とは同時には選択できない排他的選択肢として「公開したことはない」、

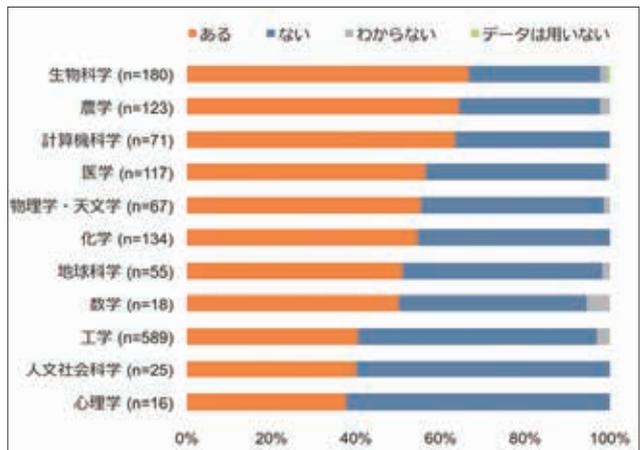
図表1 公開データとOA論文の有無（いずれもn=1,398）



図表2 データの公開方法（n=713）



図表3 分野別公開データの有無（n=1,395）



「わからない」、「研究にデジタルデータは用いない」を示した。図表1の「ある」は、公開方法を1つ以上選択した回答者の比率である。データの公開方法は「個人や研究室のサイト（50.8%）」、次いで「論文の補足資料（47.0%）」の順に選択率が高く、オープンサイエンス政策や学術雑誌のデータ共有ポリシーで推奨、あるいは想定されている永続性のあるリポジトリによる公開は「所属機関のリポジトリ」が34.2%、「特定分野のリポジトリ」が16.4%にとどまっている。

データの公開理由は、「研究成果を認知してもらいたいから（58.5%）」、「投稿した雑誌のポリシーだから（43.8%）」、「科学研究や成果実装を推進したいから（26.1%）」の順に選択率が高かった。分野別の公開率を確認すると（図表3）、生物科学（66.7%）が最も高く、同分野では「雑誌のポリシー」の選択率が最も高かった（70.8%）。データを公開していない理由は様々で、「雑誌のポリシーではないから（26.4%）」、「公開

のための時間が必要だから (25.9%)、「所属機関にポリシーがないから (22.4%)」などが比較的选择されているものの、突出した理由はみられなかった。

論文を OA にした理由は「投稿した雑誌のポリシーだったから (81.7%)」、「研究成果を認知してもらいたいから (46.9%)」に集中しており、OA にしていない理由は「雑誌のポリシーではないから (60.4%)」、「資金が必要であるから (39.6%)」の順に選択率が高かった。なお、非公開理由が解決された場合の公開意思を比較すると、データは論文よりも「いいえ」、「わからない」を選択する回答者が多く、公開に対する慎重な姿勢がうかがえた (図表 4)。

続いて、データの公開経験の有無にかかわらず、研究にデータを用いる回答者全員を対象として、“論文などの成果を発表済みの、最近の主要な研究 1 件のために収集・作成・観測したデータ (以下、「カレントデータ」)”を公開することを想定していただいた上で<sup>17)</sup>、資源の充足度や懸念の強さを尋ねた。その結果、データを整備・公開するために必要なリソースは全体的に不足しており、特に人材や時間、資金が不足していることがわかった (図表 5)。また、データ公開用のリポジリについては「わからない」とする回答が 27.1% にのぼった。分野リポジリや機関リポジリの整備が行われているものの、認知度が低い、あるいは十分ではないということが示唆された。

データを公開する場合の懸念については、「引用せずに利用される可能性」と公開データを使って「先に論文を出版される可能性」を問題視する回答者が多かった。データ公開に関する議論では、論文の出版前にデータを公開することが前提となっている場合が多いが、「先に論文を出版される可能性」に対する懸念は特に若年層で強く<sup>13)</sup>、この傾向は Tenopir ら<sup>18)</sup> や Schmidt ら<sup>19)</sup> による調査においても同様であった。公開を求めるタイミングは、慎重に検討する必要があるだろう。

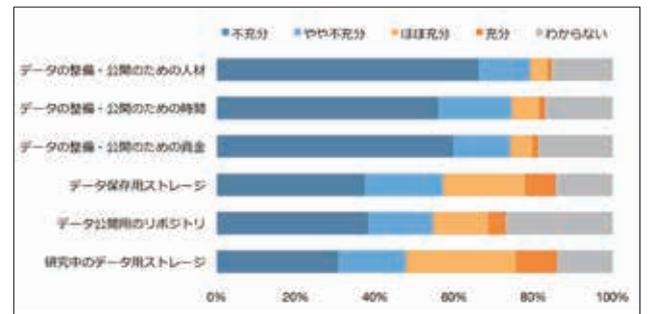
### 3. データ公開の支援に関する検討

2016 年の文部科学省『学術情報のオープン化の推進について (審議まとめ)』では、“大学等に期待される取組”として、“技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資するとともに、必要に応じて複数の大学等が共同して、データキュレーター等を育成するシステムを検討し、推進する”と述べられている<sup>20)</sup>。しかし、データの管理や公開に関するプロセスは多岐にわたるため、それぞれの難易度を尋ねることによって人材育成の参考に資することとした。具体的には、カレ

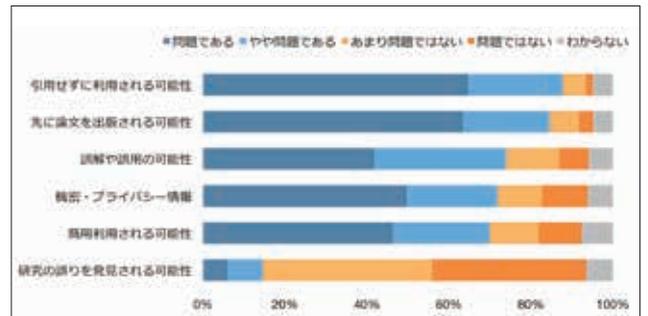
図表 4 非公開理由が解決した場合のデータと論文の公開意思 (データ n=595、論文 n=379)



図表 5 データ公開に関する資源の充足度 (n=1,396)



図表 6 データを公開する場合の懸念の強さ (n=1,396)



ントデータの整備や公開を、回答者や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目を複数選択方式で尋ねた。カレントデータが多様な場合は、最も難しいと考えられるデータについて回答していただいた。

回答者の半数以上が選択した項目は、「適切なデータ形式への変換 (60.0%)」、「データを再利用しやすいように整える (59.1%)」、「適切なメタデータ標準の選択 (50.4%)」であり、高度な専門性を有する人材の必要性が示唆された (図表 7)。一方、「機関のリポジリで公開 (26.7%)」や「異分野の研究者に紹介 (32.7%)」、「メタデータの作成 (34.2%)」は選択率が低く、専門性の必要度 (問口の広さ) という意味において、比較的支援に取り組みやすいと考えられる。

また、自由回答では、“サポートスタッフがいないと継続的に公開することは不可能と思われる”という記述もみられるなど、支援人材のニーズも確認された。反対に、“他人に行わせることはない”、“公開

図表 7 専門性を必要とする項目 (n=1,302)



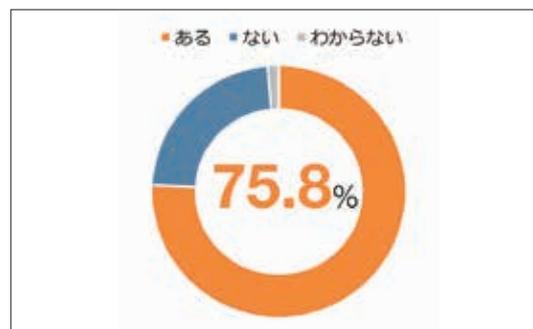
のためのデータ整理は、論文作成に近い作業であり、それを支援してもらおうというのは、論文を他人に書いてもらうのに近いことである”など、第三者による支援は不可能・困難であるとする自由回答もみられた。人的支援について検討する際には、データ公開の実担当者への聞き取りなど、より詳細な調査や議論が必要であると考えられる。

#### 4. 公開データの活用と課題

公開データの入手経験を尋ねたところ、回答者の75.8%が何らかの方法でデータを入手しており(図表8)、うち、91.2%が研究の参考に、55.3%が再利用・再分析を、46.1%が再現・追試を行っていることがわかった。冒頭で述べたように、オープンサイエンス政策は公開データを企業や市民が活用することを期待しているが、企業の研究者の70.7%が入手経験を有しており、特に研究の再現や追試を比較的良好に行っていた(54.6%)。欧州委員会(European Commission)は、データ公開が雇用や経済効果を生み出すことを想定して、欧州オープンサイエンスクラウド(European Open Science Cloud)と呼ばれる大規模なデータ公開基盤の整備を進めている<sup>21)</sup>。また、2017年6月にSpringer Nature社のデータジャーナルである*Scientific Data*は、投稿者がデータを公開する際のライセンスとしてクリエイティブ・コモンズCC0を強く推奨し、商業利用に対する制限を認めないポリシーを公開した<sup>22)</sup>。本調査では商業利用の実態までは明らかにできなかったが、今後は公開データの活用によるイノベーションの創出や経済的な利益についても注視していきたい。

データの入手に最も利用されているのは、公開と同様に「個人や研究室のサイト(64.8%)」であり、次いで「論文の補足資料(53.1%)」であった。また、データの利用においては79.8%が何らかの問題を感じており、「利用料金が必要(43.1%)」、「利用者登録が必要(33.3%)」、「利用条件がよくわからない(33.1%)」

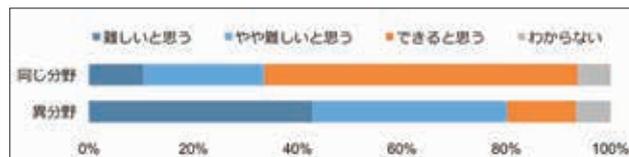
図表 8 公開データの入手経験の有無 (n=1,398)



図表 9 公開データ入手の障壁 (n=846)



図表 10 他の研究者によるデータの理解 (n=1,396)



の順に選択率が高かった(図表9)。コストをかけて公開されたデータが広く活用されるよう、リポジトリや利用条件などを整備する必要があると考えられる。

公開データの活用に関する別の課題として、データを理解することの難しさがある。他の研究者が自身のカレントデータを理解できると思うかどうか尋ねたところ、同じ分野の研究者については60.0%が「できると思う」を選択しているのに対して、異分野の研究者については13.2%にとどまった(図表10)。分野別では、人文社会科学は「できると思う」とする回答者が40.0%であったが、数学は0%、医学は4.3%と差がみられた。

研究者個人が他分野の研究者が理解できるようにデータを整備することは、その労力を考えると現実的ではない。一方で、図表6の結果からもデータを正しく解釈できないまま誤って利用されている可能性への懸念は、データ公開の障壁であることが示唆されている。公開データの信頼性を判断する方法を尋ねた質問では、著者情報(70.9%)や研究手法の確かさ(62.8%)、そのデータを用いた論文(58.2%)の順に

選択率が高かった。まずはこうした情報に容易にアクセスできるよう、データを検索するためのデータベースやメタデータを整備する必要があると考えられる。

## 5. おわりに

本調査は、研究者によるデータ公開や利用の実態と課題の一部を明らかにした。データ公開に関する課題解決の場として、2016年6月に研究データ利活用協議会<sup>23)</sup>が設立され、研究者や図書館員などの会員が活動を開始している。また、国際的な組織として研究データ同盟 (Research Data Alliance)<sup>24)</sup>があり、分野や地域を超えたデータ共有を目標に合意形成やガイドラインの策定を行っている。国内外のステークホルダーと協調することによって、効率的な課題解決を図るこ

とやグローバルスタンダードに即した研究成果の国際発信を行うことが可能になると考えられる。

さらに、データ公開は学界のみならず、政府系のオープンデータやデジタルアーカイブなど、様々な領域で盛んに進められている。こうした活動とも連携し、知見を共有することによって、互いに多くの示唆を得られるであろう。また、政策や基盤を検討する際には、多様なデータのシームレスな活用や統合を視野に入れることも重要ではないだろうか。

## 謝辞

調査及びプレテストに御協力を賜りました皆様に、心よりお礼申し上げます。

## 参考文献

- 1) G7 茨城・つくば科学技術大臣会合 (2016). つくばコミュニケ (共同声明). 内閣府. p. 9. [http://www8.cao.go.jp/cstp/kokusaiteki/g7\\_2016/2016communique.html](http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html)
- 2) 内閣府 (2016). 第5期科学技術基本計画. 53p. <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>
- 3) Foreign & Commonwealth Office (2013). G8 Science Ministers Statement. GOV.UK. 2013-6-13, <https://www.gov.uk/government/news/g8-science-ministers-statement>
- 4) 村山泰啓, 林和弘 (2014). オープンサイエンスをめぐる新しい潮流 (その1) 科学技術・学術情報共有の枠組みの国際動向と研究のオープンデータ. 科学技術動向. No. 146, p. 12-17. <http://hdl.handle.net/11035/2972>
- 5) 科学技術振興機構 (2017). オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針. 5p. [http://www.jst.go.jp/pr/intro/openscience/policy\\_openscience.pdf](http://www.jst.go.jp/pr/intro/openscience/policy_openscience.pdf)
- 6) “Over 600 Springer Nature journals commit to new data sharing policies”. Springer Nature. 2016-12-6. <http://www.springernature.com/br/group/media/press-releases/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/11111248>
- 7) Hoboken, N. J (2017). Wiley announces new Data Sharing and Citation policies to improve transparency in research. Wiley. 2017-9-14. <http://newsroom.wiley.com/press-release/all-corporate-news/wiley-announces-new-data-sharing-and-citation-policies-improve-tran>
- 8) 林和弘, 村山泰啓 (2015). オープンサイエンスをめぐる新しい潮流 (その3) 研究データ出版の動向と論文の根拠データの公開促進に向けて. 科学技術動向研究. Vol. 148, p. 4-9. <http://hdl.handle.net/11035/2999>
- 9) Clarivate Analytics 社 (旧 Thomson Reuter 社) が提供する, 論文などの学術出版物とその引用情報などを提供するデータベース。引用情報は, インパクトファクターの算出や大学ランキングにも用いられている。
- 10) Elsevier 社が提供する, 学術出版物の抄録・引用データベース。研究評価や大学ランキングなどに広く用いられている。
- 11) Beatty, Susannah (2017). Scopus makes strides in data linking. Scopus Blog. 2017-9-28. <https://blog.scopus.com/posts/scopus-makes-strides-in-data-linking>
- 12) 国際的動向を踏まえたオープンサイエンスに関する検討会 (2015). 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～サイエンスの新たな飛躍の時代の幕開け～. 内閣府. 23p. <http://www8.cao.go.jp/cstp/sonota/openscience/>
- 13) 池内有為, 林和弘, 赤池伸一 (2017). 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・学術政策研究所科学技術予測センター.
- 14) Tenopir, Carol; et al (2011). Data sharing by scientists: practices and perceptions. PLOS ONE. Vol. 6, No.6, e21101. <http://doi.org/10.1371/journal.pone.0021101>
- 15) Ferguson, Liz (2014). How and why researchers share data (and why they don't). Wiley Exchanges: our

- 
- ideas, research and discussion blog. 2014-11-3. <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>
- 16) Berghmans, Stephane; et al (2017). Open Data: The Researcher Perspective. Centre for Science and Technology Studies (CWTS). 48p. <https://www.cwts.nl/download/f-53w2.pdf>
- 17) 研究によって扱うデータの量や種類が異なる場合があると予想されるため、特定のデータ（カレントデータ）を公開することを想定して回答していただいた。
- 18) Tenopir, Carol; et al (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLOS ONE. Vol. 10, No. 8, e0134826. <http://doi.org/10.1371/journal.pone.0134826>
- 19) Schmidt, Birgit; Gemeinholzer, Birgit; Treloar, Andrew (2016). Open data in global environmental research: The Belmont Forum's Open Data Survey. PLOS ONE. Vol. 11, No. 1, e0146695. <http://doi.org/10.1371/journal.pone.0146695>
- 20) 文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会 (2016). 学術情報のオープン化の推進について (審議まとめ). 26p. [http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm)
- 21) 村山泰啓, 林和弘 (2016). 欧州オープンサイエンスクラウドに見るオープンサイエンス及び研究データ基盤政策の展望. STI Horizon. Vol. 2, No. 3, p.49-54. <http://doi.org/10.15108/stih.00044>
- 22) Open for Business. Scientific Data. 2017. <http://doi.org/10.1038/sdata.2017.58>
- 23) 余頃祐介 (2016). 集会報告「研究データ利活用協議会」公開キックオフミーティング. 情報管理. Vol. 59, No. 7, p. 490-493. <http://doi.org/10.1241/johokanri.59.490>
- 24) 白石淳子, 浅野佳那, 中島律子, 小賀坂康志 (2016). 集会報告 第8回リサーチデータ・アライアンス (RDA) 総会. 情報管理. Vol. 59, No. 9, p. 636-639. <http://doi.org/10.1241/johokanri.59.636>