

論文と研究データの公開に関する実態調査：オープンサイエンス支援 体制の構築に向けて

＜調査の趣旨＞

近年、科学研究の成果として論文や、その根拠となるデータを公開するオープンサイエンスの動きが国内外で盛んになっています。2016年5月の[G7 科学技術大臣会合](#)ではオープンサイエンスが分野横断的な課題として位置づけられ、[第5期科学技術基本計画（2016～2020年度）](#)では“国は、資金配分機関、大学等の研究機関、研究者と連携し、オープンサイエンスの推進体制を構築する”と述べられています。そして2018年6月に閣議決定された[統合イノベーション戦略](#)の「オープンサイエンスのためのデータ基盤の整備」では、主要施策として“研究分野の特性等を踏まえたオープン・アンド・クローズ戦略を考慮したデータポリシーやデータマネジメントプランの策定を促進し、これらに基づく研究データの管理・公開等を促進するとともに、公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する”とされています。

研究データの公開については、データの再利用によって研究を効率的に進めることや、異分野データを融合して産学官におけるイノベーションを創出することなどが期待されており、論文の投稿時にデータ公開を要求する学術雑誌も増加しています。一方、研究データを公開するためには人的・物的資源が必要であるため、研究のための時間や資金に限りがある産学官の皆様への支援体制の構築は喫緊の課題であると考えられます。

そこで、オープンサイエンスに関する適切な支援体制の構築に向けて、日本における現状と課題、分野や所属機関による特性、支援のニーズを明らかにするための質問紙調査を実施致します。なお、本調査は2016年に実施した[ベースライン調査](#)の後続調査という位置づけです。ベースライン調査の結果は、オープンサイエンスに関する内閣府や文部科学省の議論で活用されております。**論文や研究データを公開したご経験がないという回答も大変参考になっておりますので、ぜひ研究者の皆様からの率直なご意見をお聞かせ下さい。**

調査結果は、NISTEPの刊行物や学術出版物として公表する予定です。ご回答は個人が特定できないよう統計的に処理し、自由回答を引用する場合は匿名性を確保致します。

お忙しいところ誠に恐縮ですが、ご協力のほど何卒よろしくお願い申し上げます。

2018年10月19日
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
客員研究官 池内有為

<回答要領>

- (1) 2018年11月2日（金）までにご回答下さいますよう、お願い申し上げます。
- (2) 質問は合計32問で、回答には30分程度を要します。必須項目（*印）は19問です。
なお、選択肢で「その他」を選んだ場合は、具体的な内容の入力が必須となります。
- (3) あらかじめ質問の内容をご覧になりたい場合は、PDF形式の調査票を[こちら](#)からダウンロードできます。
- (4) ご回答は、文部科学省科学技術・学術政策研究所において厳正に管理します。
- (5) ご不明な点がある場合には、下記担当までご連絡下さい。

<留意事項>

- ・ アンケートは画面下の「回答する」をクリックして進めてください。
- ・ 前に戻る場合は、Webブラウザの「戻る」（「←」ボタン等）で戻ってください。
- ・ 回答を一時的に保存したり、再開する機能をご用意していません。途中で回答を中止する際には、再度最初の質問から回答し直して頂く必要があります。
- ・ 回答画面（回答用URL）にはタイムアウトによるセッションの切断がないため、ブラウザを閉じない限りは、回答状況が維持されます。
- ・ 回答中にページを再読み込み（リロード）した場合、アンケート開始ページに遷移し、それまでの回答は破棄されますのでご注意ください。
- ・ 回答データは、「送信」を押さない限り送信されないため、回答終了後、忘れずに「送信」ボタンを押してください。なお、「送信」後の修正は出来ません。

以上の留意事項をふまえて、ご協力をよろしくお願い致します。

<アンケート担当>

科学技術予測センター

田中綾子 (stfc@nistep.go.jp)

本調査は、オープンサイエンス（学術論文や、その根拠となる研究データを公開すること）についておうかがいします。

1. 研究分野について

Q1 ご自身の研究分野に最も近いものをお選び下さい。*

- | | | |
|----------------------------------|------------------|----------|
| (1)工学 | (2)天文学 | (3)化学 |
| (4)物理学 | (5)地球科学 | (6)数学 |
| (7)コンピュータサイエンス | (8)農学 | (9)生物科学 |
| (10)医学 | (11)心理学 | (12)社会科学 |
| (13)人文学 | (14)その他 [具体的に：] | |
| (15)論文出版や学会等での口頭発表は行っていない【→Q32へ】 | | |

2. 学術論文について

ご自身の論文のオープンアクセス（公開）状況や、論文の利用についておうかがいします。

なお、本調査における「オープンアクセス」とは、論文がインターネットで公開され、読者は無料で読むことができる状態とします。

【オープンアクセスの例】

- ・ オープンアクセスの雑誌で出版する（PLOS など）
- ・ 雑誌のオープンアクセスオプションを選択する（Springer Open Choice など）
- ・ 雑誌等が一定期間経過後に論文をオープンアクセスにする
- ・ セルフアーカイブ：機関リポジトリやプレプリントサーバで論文を公開する（arXiv など）

Q2 ご自身の論文について、あてはまるものをお選び下さい。*

- (1) オープンアクセスの論文がある
- (2) オープンアクセスの論文はない【→Q4へ】
- (3) わからない【→Q6へ】

Q3 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。[複数選択可]*

- (1) 他の研究者からのリクエストに応じて
 - (2) 論文を投稿した雑誌がオープンアクセスだから
 - (3) 助成機関のポリシー（助成条件）だから
 - (4) 所属機関のポリシーだから
 - (5) 分野・コミュニティの規範だから
 - (6) 研究成果を広く認知してもらいたいから
 - (7) 科学研究や成果実装を推進したいから
 - (8) オープンアクセスに貢献したいから
 - (9) その他 [具体的に：]
 - (10) あてはまるものはない【※黄色の回答は他の選択肢とは同時に選べません】
- 【すべて→Q6へ】

Q4 論文をオープンアクセスとしていない理由として、あてはまるものをお選び下さい。[複数選択可] *

- (1) 資金がないから
- (2) 時間がないから
- (3) ニーズがないと思うから
- (4) 投稿したい雑誌がオープンアクセスではないから
- (5) 助成機関のポリシー（助成条件）ではないから
- (6) 所属機関にオープンアクセス方針がないから
- (7) 分野・コミュニティで推奨されていないから
- (8) リポジトリなどの公開手段がないから
- (9) その他 [具体的に：]
- (10) あてはまるものはない【→Q6へ】

Q5 Q4の理由が解決された場合、論文をオープンアクセスにしたいと思われますか？ *

- (1) はい
- (2) いいえ
- (3) わからない

Q6 論文やプレプリントをオープンアクセスにしようとする場合、次の点はどの程度問題となりますか？ あてはまるものをそれぞれお選び下さい。 *

	問題ではない	あまり問題ではない	やや問題である	問題である	わからない
商用利用される可能性	1	2	3	4	0
機密・プライバシー情報	1	2	3	4	0
著作権	1	2	3	4	0
誤解や誤用の可能性	1	2	3	4	0
不正利用・改ざんの可能性	1	2	3	4	0
二次利用に関して責任が生じる可能性	1	2	3	4	0
引用せずに利用される可能性	1	2	3	4	0
研究の誤りを発見される可能性	1	2	3	4	0
公開したプレプリント*を使って自分より先に論文を出版される可能性	1	2	3	4	0

* 学術雑誌などで出版される前の原稿

Q7 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

- (1) サーチエンジン (Google, Google Scholar など)
- (2) 論文情報のデータベース (Web of Science, Scopus, CiNii Articles など)
- (3) 出版社や学術雑誌のサイト (Elsevier, Wiley など)
- (4) 学術機関のリポジトリ・アーカイブ (大学やNASAの機関リポジトリなど)
- (5) 論文やプレプリントのサーバ (PubMed Central, arXiv, J-Stage など)
- (6) 論文や学術記事の参考文献
- (7) 政府・機関・出版社などの広報, ニュースレター
- (8) ブログや一般的な SNS (Facebook, Twitter など)
- (9) 学術系 SNS (Mendeley, ResearchGate など)
- (10) アラートサービス (RSS など)
- (11) メーリングリスト
- (12) 研究者や同僚に尋ねる／教えてもらう
- (13) その他 [具体的に: _____]

3. 研究データの提供について

本調査における「研究データ」とは、研究のために収集・作成・観測したデジタルデータを指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しません。

データの「提供」とは、E-mail や USB フラッシュメモリ、クラウドサービス (Dropbox や Google Drive) などを使って、共同研究者を除く他者に渡す (共有する) ことを指します。特定の人以外はアクセスできない状態です。

Q8 共同研究者を除く他の研究者に研究データを提供したご経験はありますか？

- (1) よくある
- (2) たまにある
- (3) ほとんどない
- (4) まったくない 【→Q10へ】
- (5) わからない 【→Q10へ】

Q9 共同研究者を除く他の研究者に研究データを提供した方法をお選びください。[複数回答可]

- (1) E-mail
- (2) 外部記憶媒体 (USB フラッシュメモリ, 記録型 CD/DVD など)
- (3) 所属機関のクラウドサービス
- (4) 外部のクラウドサービス (Dropbox, Google Drive など)
- (5) その他 [具体的に: _____]
- (6) わからない

4. 公開データの利用について

「公開データ」とは、ウェブサイトやリポジトリ、論文の補足資料などに掲載され、インターネットでアクセスして利用できるデータを指します。利用料金や利用者登録が必要な場合も含まれます。

Q10 これまでに、公開データを以下の公開先から入手したご経験はありますか？ [複数選択可] *

- (1) 個人や研究室のウェブサイト
- (2) 学術機関のリポジトリ・データアーカイブ（大学やNASAのリポジトリなど）
- (3) 特定分野のリポジトリ・データアーカイブ（DDBJやICPSRなど）
- (4) データ共有サービス（figshare, zenodo など）
- (5) コード共有サービス（GitHub など）
- (6) 論文の補足資料
- (7) 学術系 SNS（Mendeley, ResearchGate など）
- (8) その他 [具体的に：]
- (9) ない【→Q14へ】
- (10) わからない【→Q14へ】

Q11 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。 [複数選択可] *

- (1) 利用料金が必要
- (2) 利用者登録が必要
- (3) リクエストから入手までに時間がかかる
- (4) 最新のデータを入手できない
- (5) データごとにフォーマットが異なる
- (6) データごとに品質が異なる
- (7) データの解釈や利用方法がよくわからない
- (8) 利用したいデータへのアクセス方法がよくわからない
- (9) 利用条件（営利利用が可能かどうかなど）がよくわからない
- (10) 著作者情報がよくわからない
- (11) その他 [具体的に：]
- (12) 問題を感じたことはない

Q12 入手した公開データの利用目的について、あてはまるものをお選び下さい。 [複数選択可]

- (1) 自身の研究のアイデアや仮説の参考にする
- (2) 再分析・再利用して自身の研究を行う
- (3) 研究を再現・追試する
- (4) その他 [具体的に：]
- (5) あてはまるものはない

Q13 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

- (1) サーチエンジン (Google, Google Dataset Search など)
- (2) データ情報のデータベース (Data Citation Index など)
- (3) 学術機関のリポジトリ・データアーカイブ (大学や NASA のリポジトリなど)
- (4) 出版社や学術雑誌のサイト (Elsevier, Wiley など)
- (5) 特定のデータリポジトリ [具体的に：]
- (6) 論文や学術記事の参考文献
- (7) データジャーナル (簡易なデータ記述とデータへのリンクを掲載した雑誌)
- (8) 政府・機関・出版社などの広報, ニュースレター
- (9) ブログや一般的な SNS (Facebook, Twitter など)
- (10) 学術系 SNS (Mendeley, ResearchGate など)
- (11) アラートサービス (RSS など)
- (12) メーリングリスト
- (13) 研究者や同僚に尋ねる／教えてもらう
- (14) その他 [具体的に：]

5. データマネジメントプラン (DMP) について

データマネジメントプラン (DMP) についておうかがいします。

本調査における「DMP」とは、研究データを適切に管理・公開・保存することを目的として、研究開始から研究終了後までのデータの取り扱いについて記載する文書を指します。助成機関や研究機関が研究費獲得の際に提出を求める場合があります。

【DMP の記載内容の例】

- ・ データの概要 (データの種類, 量など)
- ・ データの公開・非公開の方針 (一般公開／共同研究者のみ／非公開, 公開時期など)
- ・ データの保存場所 (研究中・研究後)

Q14 これまでに、DMP を作成したご経験はありますか？ [複数選択可] *

- (1) 科学技術振興機構 (JST) による助成研究の DMP
- (2) 日本医療研究開発機構 (AMED) による助成研究の DMP
- (3) 新エネルギー・産業技術総合開発機構 (NEDO) による助成研究の DMP
- (4) 所属機関の DMP
- (5) 個人や研究グループのための DMP
- (6) その他 [具体的に：]
- (7) ない【→Q16 へ】
- (8) わからない【→Q16 へ】

Q15 DMP を作成した理由として、あてはまるものをお選び下さい。[複数選択可] *

- (1) 研究を行う際の習慣だから
- (2) 効率的にデータを管理するため
- (3) 助成機関が要求しているから
- (4) 所属機関が要求しているから
- (5) その他 [具体的に：]
- (6) あてはまるものはない
【すべて→Q17へ】

Q16 DMP を作成していない理由として、あてはまるものをお選び下さい。[複数選択可] *

- (1) DMP を知らなかったから
- (2) DMP の必要性を感じないから
- (3) DMP を作成する時間がないから
- (4) 助成機関から要求されていないから
- (5) 所属機関から要求されていないから
- (6) その他 [具体的に：]
- (7) あてはまるものはない

6. 研究データの公開について

これまでに、ご自身の研究データをインターネットで公開したご経験についておうかがいします。

(再掲) 本調査における「研究データ」とは、研究のために収集・作成・観測したデジタルデータを指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しません。

Q17 これまでに、研究データを以下の方法で公開したご経験はありますか？[複数選択可]

*

- (1) 個人や研究室のウェブサイト
- (2) 所属機関のリポジトリ・データアーカイブ
- (3) 特定分野のリポジトリ・データアーカイブ (DDBJ や ICPSR など)
- (4) データ共有サービス (figshare, zenodo など)
- (5) コード共有サービス (GitHub など)
- (6) 論文の補足資料
- (7) 学術系 SNS (Mendeley, ResearchGate など)
- (8) その他 [具体的に：]
- (9) 公開したことはない【→Q21へ】
- (10) わからない【→Q23へ】
- (11) 研究にデジタルデータは用いない【→Q32へ】

Q18 研究データを公開した理由として、あてはまるものをお選び下さい。[複数選択可] *

- (1) 他の研究者からのリクエストに応じて

- (2) 論文を投稿した雑誌のポリシー（投稿規定）だから
- (3) 助成機関のポリシー（助成条件）だから
- (4) 所属機関のポリシーだから
- (5) 分野・コミュニティの規範だから
- (6) 業績になる場合があるから
- (7) 研究成果を広く認知してもらいたいから
- (8) 科学研究や成果実装を推進したいから
- (9) オープンデータに貢献したいから
- (10) その他 [具体的に：]
- (11) あてはまるものはない

Q19 研究データの公開によって良い結果が得られたご経験がありましたら、差し支えない範囲でお書き下さい。（最大 500 文字） [例：共同研究につながった]

Q20 研究データの公開によって問題が起きたご経験がありましたら、差し支えない範囲でお書き下さい。（最大 500 文字）

【すべて→Q23 へ】

Q21 研究データを公開していない理由として、あてはまるものをお選び下さい。 [複数選択可] *

- (1) 資金がないから
- (2) 時間がないから
- (3) ニーズがないと思うから
- (4) 論文を投稿した雑誌のポリシー（投稿条件）ではないから
- (5) 助成機関のポリシー（助成条件）ではないから
- (6) 所属機関にポリシーがないから
- (7) 分野・コミュニティで推奨されていないから
- (8) 業績にならないから
- (9) リポジトリなどの公開手段がないから
- (10) その他 [具体的に：]
- (11) あてはまるものはない 【→Q23 へ】

Q22 Q21 の理由が解決された場合、研究データを公開したいと思いますか？ *

- (1) はい
- (2) いいえ

(3) わからない

Q23 研究データ公開のインセンティブとして、次の点はどの程度重要だと思われますか？
あてはまるものをそれぞれお選び下さい。*

	重要ではない	あまり重要ではない	やや重要	重要	わからない
利用者からの報告	1	2	3	4	0
利用料金などの対価	1	2	3	4	0
データに紐づいた論文の引用	1	2	3	4	0
データの引用*	1	2	3	4	0
研究資金の獲得	1	2	3	4	0
業績として論文と同様に評価されること	1	2	3	4	0

*論文と同様に、参考文献リストにデータ作成者やデータ名、識別子などを記載する

7. 最近の研究のためのデータについて

ここからは、研究データの整備や公開に必要な支援について検討するための質問です。

研究ごとに扱うデータの種類や量が異なる場合が多いと考えられるため、「論文などの成果を发表済みの、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」についておうかがいします（以下、「カレントデータ」と記します）。

データ公開のご経験がない場合は、カレントデータのうち、論文等の根拠となる部分を公開すると想定してお答え下さい。

【例：研究1件のために質問紙調査とインタビュー調査を実施した場合】

<カレントデータ>

- ・質問紙の回答を入力したスプレッドシート
- ・回答を分析するためのRのコード
- ・インタビューを録音した音声データ
- ・インタビューを書き起こしたテキストデータなど

<データ公開>

質問紙調査の回答データを第三者が再利用できるように、調査概要の説明を作成、回答を入力したスプレッドシートから個人情報削除、項目に見出しをつけるといった処理を行った上で、質問紙や分析のためのコードとともにリポジトリに登録する。

Q24 カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。

(1) MB（メガバイト）以下

- (2) GB (ギガバイト)
- (3) TB (テラバイト)
- (4) PB (ペタバイト) 以上
- (5) わからない

Q25 現在までに、カレントデータ（データの一部）を公開しましたか？*

- (1) はい：以降の質問は、(カッコ内)をお読みの上でお答え下さい
- (2) いいえ：以降の質問は、公開することを想定してお答え下さい
- (3) わからない：以降の質問は、公開することを想定してお答え下さい

Q26 カレントデータを整備・公開しようとする場合、次の資源は十分に整っていますか（いましたか）？ あてはまるものをそれぞれお選び下さい。*

	不十分	やや不十分	ほぼ十分	十分	わからない
研究中のデータ用ストレージ	1	2	3	4	0
データ公開用のリポジトリ	1	2	3	4	0
研究終了後のデータ保存用ストレージ	1	2	3	4	0
データの整備・公開のための時間	1	2	3	4	0
データの整備・公開のための人材	1	2	3	4	0
データの整備・公開のための資金	1	2	3	4	0

Q27 カレントデータを公開しようとする場合、次の点はどの程度問題となりますか（なりましたか）？ あてはまるものをそれぞれお選び下さい。*

	問題ではない	あまり問題ではない	やや問題がある	問題がある	わからない
商用利用される可能性	1	2	3	4	0
機密・プライバシー情報	1	2	3	4	0
データの所有権・契約	1	2	3	4	0
誤解や誤用の可能性	1	2	3	4	0
不正利用・改ざんの可能性	1	2	3	4	0
二次利用に関して責任が生じる可能性	1	2	3	4	0
引用せずに利用される可能性	1	2	3	4	0
研究の誤りを発見される可能性	1	2	3	4	0
公開したデータを使って自分より先に論文を出版される可能性	1	2	3	4	0

Q28 カレントデータを公開した場合、公開したカレントデータは、ご自身と同じ分野、すなわち Q1 で選択した分野の研究者の多くが理解できると思われませんか？

- (1) できると思う：データにつけた説明（Readme ファイルなど）を読めば理解できると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q29 カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われませんか？

- (1) できると思う：データにつけた説明（Readme ファイルなど）を読めば理解できると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q30 カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい [複数選択可]

- (1) 適切なデータ形式
- (2) 適切なリポジトリ
- (3) 適切なメタデータ（作成者やキーワードなど、データを検索するために付与するデータの要約情報）標準（例：ジャパンリンクセンターによる[研究データのメタデータ仕様 \(PDF\)](#)）
- (4) データのバージョン管理方法
- (5) データのバックアップ方法
- (6) データの安全な管理方法（セキュリティ）
- (7) データに含まれる機密情報の処理
- (8) 知的財産権やライセンス
- (9) その他 [具体的に： _____]
- (10) あてはまるものはない

Q31 カレントデータの整備や公開を、ご自身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お選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。[複数選択可]

- (1) 適切なデータ形式への変換
- (2) 適切なリポジトリの選択
- (3) 適切なメタデータ（作成者やキーワードなど、データの要約情報）標準の選択
- (4) メタデータの作成
- (5) データを再利用しやすいように整える
- (6) 機関のリポジトリによるデータ公開
- (7) データを異分野の研究者に紹介する
- (8) その他 [具体的に： _____]
- (9) あてはまるものはない（いずれも専門性を必要としない）

8. 自由回答

Q32 論文やデータの公開，オープンサイエンス，および調査に関するご意見や感想がありましたら，ご自由にお書き下さい。（最大 1,000 文字）

ご協力ありがとうございました。