厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）
総括研究報告書

厚生労働分野のオープンサイエンス推進に向けたデータポリシー策定に資する研究
オープンサイエンスに関する調査アンケート報告

木村　映善　国立保健医療科学院保健医療情報管理分野 統括研究官

研究要旨

統合イノベーション戦略においてオープンサイエンスを推進するための社会インフラとしてのデータ基盤の整備が明示され、国立研究開発法人はデータポリシーを 2020 年度中に策定し、これに基づく研究データの管理・公開等を促進することとなった。また、データポリシーの策定にあたって「国立研究開発法人におけるデータポリシー策定のためのガイドライン」（平成 30 年 6 月 29 日）を参考にすることが求められている。これを受けて、データポリシーを策定する厚生労働省が所管する研究機関で策定すべきデータポリシーに関して、機関共通で取り組むべき事項、各研究機関の特性に応じて取り組むべき事項を整理し、厚生労働行政等に資する研究データの利活用を最大限に促進するために必要な、一貫性及び整合性のあるデータポリシーの要件を明らかにするとともに、機関リポジトリの整備及び運用に関する提言を取りまとめることを目的として、研究機関に所属する研究者に対してアンケート調査を実施した。調査結果からは、データ公開に対して一定の理解を示しつつも、データ公開という要求事項の増加による研究業務への圧迫や機微な個人情報を含むデータの扱いへの懸念がみられた。本調査の結果をもとに、必要な人材像、研究機関、国がそれぞれに優先的に取り組むべき課題について整理し、報告する。

キーワード　オープンサイエンス、オープンデータ、データポリシー

水島 洋（国立保健医療科学院　研究情報支援研究センター　センター長）
星　佳芳（国立保健医療科学院　研究情報支援研究センター　上席主任研究官）
上野　悟（国立保健医療科学院　研究情報支援研究センター　主任研究官）
日野原　友佳子（国立研究開発法人医薬基盤・健康・栄養研究所　研究企画評価主幹）
堀内　直哉（国立研究開発法人医薬基盤・健康・栄養研究所・戦略企画部・部長）
青柳　一彦（国立研究開発法人国立がん研究センター　研究支援センター・リサーチ・アドミニストレーター）
岩上　直嗣（国立研究開発法人国立循環器病研センター　研究振興部　研究医療課・課長代理）
波多野　賢二（国立研究開発法人国立精神・神経医療研究センター　トランスレーショナル・メディカル・センター・室長）
石井　雅通（国立研究開発法人国立国際医療研究センター 医療情報基盤センター 副センター長）
野口　貴史（国立研究開発法人国立成育医療研究センター　情報管理部・部長）
滝川　修（国立研究開発法人国立長寿医療研究センター　治験・臨床研究推進センター開発・連携推進部・部長）
斎藤　嘉朗（国立医薬品食品衛生研究所 医薬安全科学部・部長）
小島　克久（国立社会保障・人口問題研究所 情報調査分析部・情報調査分析部長）
大澤　英司（国立感染症研究所　企画調整

主幹）
小野　栄一（国立障害者リハビリテーションセンター　研究所長）
小林　卓馬（独立行政法人国立病院機構本部総合研究センター・治験専門職）
髙橋　洋（独立行政法人労働者健康安全機構　研究試験企画調整部・部長）
日詰　正文（独立行政法人国立重度知的障害者総合施設のぞみの園　研究部・部長）

A.　研究目的

A.1背景

統合イノベーション戦略（平成 30 年 6 月 15 日閣議決定）において「オープンサイエンスのためのデータ基盤の整備」、つまり、すべての者がサイバー空間の研究データを利活用し、協働によってイノベーションを創出するというオープンサイエンスを推進するための社会インフラとしてのデータ基盤の整備が明示され、国立研究開発法人（以下、国研）は、研究分野の特性、国際的環境、産業育成等に配慮したデータポリシーを 2020 年度中に策定し、これに基づく研究データの管理・公開等を促進することとなった。また、データポリシーの策定にあたっては、内閣府総合科学技術・イノベーション会議（CSTI）が設置した「国際的動向を踏まえたオープンサイエンスの推進に関する検討会」による「国立研究開発法人におけるデータポリシー策定のためのガイドライン」（平成 30 年 6 月 29 日）を参考にすることが求められている。

これを受けて厚生労働省は、厚生科学審議会科学技術部会（平成 30 年 7 月開催）での議論を踏まえ、厚生労働省に属する研究機関においてもデータポリシーを策定し、所管する全ての研究機関でオープンサイエンスを推進することとなった。

上記の「ガイドライン」によれば、データポリシーは研究機関等の研究分野の特性やミッション等に基づいて定められるものとされているが、厚生労働行政や他の分野の政策等に資するデータの利活用を促進するためには、研究機関等で一定の共通する事項や内容を設定して、研究データの横断的連携の推進等に向けて相互運用性を高める必要もある。しかしデータポリシーの記載事項や記載内容をどこまで機関共通にすべきか、その具体的な考え方や根拠は明らかにされていない。また、研究データの公開・利活用の基盤となる機関リポジトリに関しても、具体的な整備・運用の方法が確立していない。

そこで本研究では、厚生労働省が所管する研究機関で策定すべきデータポリシーに関して、機関共通で取り組むべき事項、各研究機関の特性に応じて取り組むべき事項を整理し、厚生労働行政等に資する研究データの利活用を最大限に促進するために必要な、一貫性及び整合性のあるデータポリシーの要件を明らかにするとともに、機関リポジトリの整備及び運用に関する提言を取りまとめることを目的とする。

B.　研究方法

2.1 調査対象

2019 年 11 月から、各研究機関内部の倫理委員会の承認が得られた研究機関からアンケート調査に参加頂いた。厚生労働省所管の研究機関である、医薬基盤・健康・栄養研究所、国立がん研究センター、国立循環器病研究センター、国立精神・神経医療研究センター、国立国際医療研究センター、国立成育医療研究センター、国立長寿医療研究センター、国立医薬品食品衛生研究所、国立保健医療科学院、国立社会保障・人口問題研究所、国立感染症研究所、国立障害者リハビリテーションセンター、独立行政法人国立病院機構、独立行政法人労働者健康安全機構、独立行政法人国立重度知的障害者総合施設のぞみの園、独立行政法人労働政策研究・研修機構から１名ずつ各分担研究者として参画いただき、各分担研究者から所内メーリングリスト及び文面にて、研究機関の規模に応じて 2 週間前後の回答期間を設定し、調査への協力依頼を送信した。

本研究の主眼は厚労省所管の研究機関に

おける保有データの状況の網羅的把握と、先行研究における保健医療分野に関する委細の検討に乏しい状況にあってより委細な状況把握に努めることにある。分担研究者において、アンケート回収率を高めるために厚生労働省及び研究班からの依頼状を添え、組織内の各分野の長及び所属研究員への一斉同報による二度以上のアンケート回答依頼を行った。また、今回の調査の目標として網羅性の確保に重点を置いたため、回答数もそうであるが、各分野に広く行き渡ることが重要であるため、各部門から最低１名以上の回収をしていただくように働きかけた。

また保有データの定義については、基本的にオープンデータに直ちに加工・公開できる可能性があるデジタルデータ、あるいはデジタル化されたデータ（紙資料であれはスキャンしてPDF化したもの等）とし、物理的な試料（生物由来物、鉱物）等は含まないこととした。但し、試料から採取された測定、検査結果、3D スキャンした構造データ等は対象に含むこととした。 社会科学系の研究等においては、インタビュー調査のメモ等、紙媒体で保存管理しているものがあるが、今回はそこまでは求めず、スキャン等されてデジタル化されているものを対象とした。

B.2質問項目の設計方針

これまでのデータ公開に関する質問調査はSTEM分野、生物学分野、自然科学分野、多分野等において実施されてきたが、医療分野の研究者集団において実施したものは発表されていない。先行研究と比較を行い、医療分野特有の課題を抽出するには先行研究の質問票と問題を揃えることが望まれる。しかし、全てを集約すると膨大な質問数になり、分担研究者によるプレテストを実施したところ、回答時間に30分を越え、回答者が疲労することが指摘された。そこで、本研究の契機であるオープンデータポリシーの策定に資するための研究者と保有データの現状把握と公開に関する障壁を明らかにする設問に絞ることとした。オープンサイエンス、オープンデータについて馴染みが薄い研究者がいることが先行研究から予想されたことから、質問票の冒頭にオープンサイエンスの定義について説明した。

B.3質問の構成

本節で参照する括弧内の設問番号は別添の質問票内の設問番号に対応する。

B.3.1　公開データの利用状況

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(F1)については、池内ら[1]の調査と同じ設問とした。公開データの入手先(F2)については、池内ら[1]の調査をベースに、論文補足資料について、「出版社」「論文データベース」「学術データアーカイブ」とより入手先を細分化し、リポジトリについて医学分野で知られているリポジトリを例示した。

B.3.2　データ保持状況

研究データを保持するモチベーション(G1)及び研究データの保持に関する課題(G3)、データ保持に関する知識を増やす有用な方法(G5)では、Kuipers[2]らの質問票と同じ項目としたが、選択肢の数はLikert尺度[3]として5段階に設定した。なお、設問G3については、参考にした調査票において、「Threats to digital preservation」と「General threats to currently preserved digital data」のそれぞれの設問の選択肢を統合したものとした。さらに設問G2には、外的要因として、「国による研究データに関する政策、規則の一貫性の欠如」、「所属研究機関による研究データに関する規則・方針の不整備」を追加した。

B.3.3　データの保持対象

研究遂行上でよく使うデータのフォーマット(H1)について、Kuipers[2]らの質問票の項目に加えて医療分野に特化した項目として「医用画像（DICOM等）」を追加した。所謂医療標準規格であるHL7メッセージや診療情報交換規約文書（CDA: Clinical Document Architecture)は、構造化されたテキスト・データの項目として例示した。研

究者が所属している研究分野(H2)については、我が国での研究分野の分類に準拠することで、オープンデータに関するポリシー策定の参考となることを意図して、「e-Rad における研究分野一覧[4]」を採用した。研究者によっては複数領域での活動がみられるため、最大 5 個までの申告をできるようにした。研究でよくつかうデータの分野(H3)について、Tenopir ら[5]の質問票の項目に加えて、厚労省所管の研究機関における研究で利用されている「政府統計、国際機関の統計」「行政記録・行政資料 」「文献資料（古記録、新聞、書籍)」を追加した。

研究データの対象物(H4)は本調査で独自に定義した項目であり、機微な個人情報を含む可能性があるもの、データとして公開可能なものであるかを特定するために、分担研究者でのブレインストーミングによって 20 項目を決定した。

B.3.4 学際的な研究データの利用

データの公開状況(I1)について、Kuipers[2]らの質問票の項目に加え、我が国におけるデータ提供の実態を踏まえ、「データは法律などの条件を満たす場合に限り公開している」「契約を締結した上でデータを無償で公開している」「その他」の 3 項目を追加した。データ公開に関する障壁(I2)は、Kuipers[2]らの質問票の項目と同じものにした。

B.3.5 研究データの提供状況

データ提供形態(J1)について、Tenopir ら[5]の質問票の項目に加えて、Ferguson らの調査[6]から「学会や助成機関の Web サイト」「一般公開の為のリポジトリ利用」の項目を追加し、本調査独自項目として「CD や USB 等の物理媒体で直接 」を追加した。また、「研究者間に閉じた環境」について「組織間ネットワークや共同研究者のみに開示されたサーバ等、一定の加入手続き等を経ないとアクセスできないような制限を設けている配布形態はこのカテゴリに入ります。他の所は、基本的にアクセス制限がない状態での公開を想定しています。」と定義した。研究データに関するプロセスへの満足状況(J3)について Tenopir[5]の質問票の項目に加えて、オープンデータの公開準備に必要なプロセスとして、「メタデータを準備するツール」「データに関する文書（資料）を準備するためのツール」の 2 つを追加した。研究者が所属している組織・プロジェクト(J4)について Tenopir[5]の質問票と同じ項目を採用した。データ利用に関する懸念(J5)について、Tenopir[5]、小野[7]の質問票の項目群から、データ利用に関する懸念につながる項目を選択した。データの裁量に関する条件(J7)について Tenopir[5]の質問票の項目を採用した。異分野の研究での再利用可能性(J8)について、池内[1]らの調査結果にもとづいて、データの説明、互換性、ニーズの影響について問う設問を作成した。他者がデータを利用する際の公正な条件(J10)について、Tenopir[5]、小野[7]の質問票の項目群から採用した。データを公開する際のライセンス形態(J11)は小野[7]の質問票の項目を採用した。データを公開した時の関心のある指標(J13)については研究班によって検討した。第三者に提供できない理由(J14)について、小野[7]の質問票をベースに、権利・法律・組織についてひとまとめにしていた選択肢を個別の選択肢に細分化し、Tenopir[5]から「研究助成者によってデータ公開を要求されていない」を追加した。また、この設問に伴い、提供できない理由が解決された場合に研究データを公開できるかを問う質問(J15)を追加した。DMP の提出を求めた助成機関について問う質問は Tenopir[5]を参考に、厚労省管轄下の研究機関が主に利用している助成機関の一覧を作成した。

B.3.6 データに関する技能

データ公開にあたり、今後受けたいトレーニング等(K1)について、池内ら[1]の質問票を採用した。専門性のある第三者による支援が必要な物(K2)として、池内ら[1]の質問票の項目に、「匿名加工の実施」「知的財産権やライセンス」を追加した。データ公開

に関わった研究者の業績のための指標について研究班で項目を検討した。

B.4分析方法

Web アンケートシステムは国立保健医療科学院がホストしている仮想環境上にオープンソースの Lime Survey[8]というソフトウェアを用いて構築した。アンケート終了後、Lime Survey からデータ抽出し、統計処理ソフトウェア R[9]にて分析した。統計分析において有意水準 p は 5%とし、適宜信頼区間を提示する。各統計処理については、結果ごとに適用した統計処理を提示する。回答に複数選択肢があるものはリッカート尺度[3]にもとづき、Likert パッケージ[10]を利用して分析した。

C. 研究結果

C.1 回答者の属性

本研究は紙媒体の調査票を研究者個人に送信する方法ではなく、各研究機関内の研究者向けの連絡メーリングリスト、個別依頼などを通して Web アンケートシステム上での回答を依頼したものであり、厳密に対象回答者数が事前に特定されているわけではない。事前に 14 機関の分担研究者より申告されていた本アンケートの回答対象となる研究者数の総計は 2589 人であった。

14 機関全体における最終有効回答者数は 407 名(回答率 15.72%)であった（厚労省管轄外の組織からの回答、所属情報未記入の回答は除外)。これらの回答を分析対象とした（表 1)。回答者の所属と年齢層別の集計結果を表 2 に示す。また、研究年数、年齢、性別の相関図を図 1 に示す。相関図において、経験年数(Experience)は５年未満、5-10 年、11-15 年、16-20 年、20 年以上の５階級、年齢(Age)は 30 歳未満、30-34 歳、35-39 歳、40-44 歳、45-49 歳、50-54 歳、55-59 歳、60 歳以上の 8 階級である。年齢は、男性は 45-49 歳、女性は 40-44 歳と、女性の方が若い方に最頻値が分布している。経験年数は、男女ともに経験年数が長い人間の方が多い傾向があり、年齢と経験の関係において、相関関係（ポリコリック相関係数:-0.7916、$p<0.05$)が観察された。すなわち、厚労省所管の研究機関は全体的に中年期の研究者が多く、研究期間が長い者が多い傾向にある。他の研究機関等で実績を積んでから機関の研究者として着任するというキャリアの積み方が多いことが伺える。

表 1 アンケート参加機関と回答者数

No	研究機関名	回答者数
1	国立保健医療科学院	21
2	国立医薬品食品衛生研究所	65
3	国立感染症研究所	22
4	国立研究開発法人医薬基盤・健康・栄養研究所	24
5	国立研究開発法人国立がん研究センター	112
6	国立研究開発法人国立国際医療研究センター	20
7	国立研究開発法人国立循環器病研究センター	7
8	国立研究開発法人国立精神・神経医療研究センター	54
9	国立研究開発法人国立長寿医療研究センター	9
10	国立社会保障・人口問題研究所	10
11	国立障害者リハビリテーションセンター	31
12	独立行政法人労働者健康安全機構	16
13	独立行政法人国立病院機構	10
14	独立行政法人国立重度知的障害者総合施設のぞみの園	6

C.2 公開データの利用状況

本節では公開データの利用につながる行動の特定の参考にするため、公開データを探す時に頻用するツール及び入手先について問うた結果の集計と自由回答を示す。設問の番号は質問票における番号と対応する。

C.2.1 公開データを探す時によく使用するツール

公開データの検索方法を確認するため、「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であてはまるものをすべてお選び下さい(Q F1)」と尋ねた結果を表 3,4 に示す。質問では複数選択肢の他、自由記述欄を提供した。

表 2 回答者の年齢と性別

Age	*Sex* 男性	女性	***Total***
30歳未満	4 1 %	1 0.2 %	5 1.2 %
30-34歳	16 3.9 %	5 1.2 %	21 5.1 %
35-39歳	36 8.8 %	14 3.4 %	50 12.2 %
40-44歳	52 12.8 %	22 5.4 %	74 18.2 %
45-49歳	73 17.9 %	20 4.9 %	93 22.8 %
50-54歳	58 14.3 %	19 4.7 %	77 19 %
55-59歳	48 11.8 %	16 3.9 %	64 15.7 %
60歳以上	19 4.7 %	4 1 %	23 5.7 %
Total	306 75.2 %	101 24.8 %	407 100 %

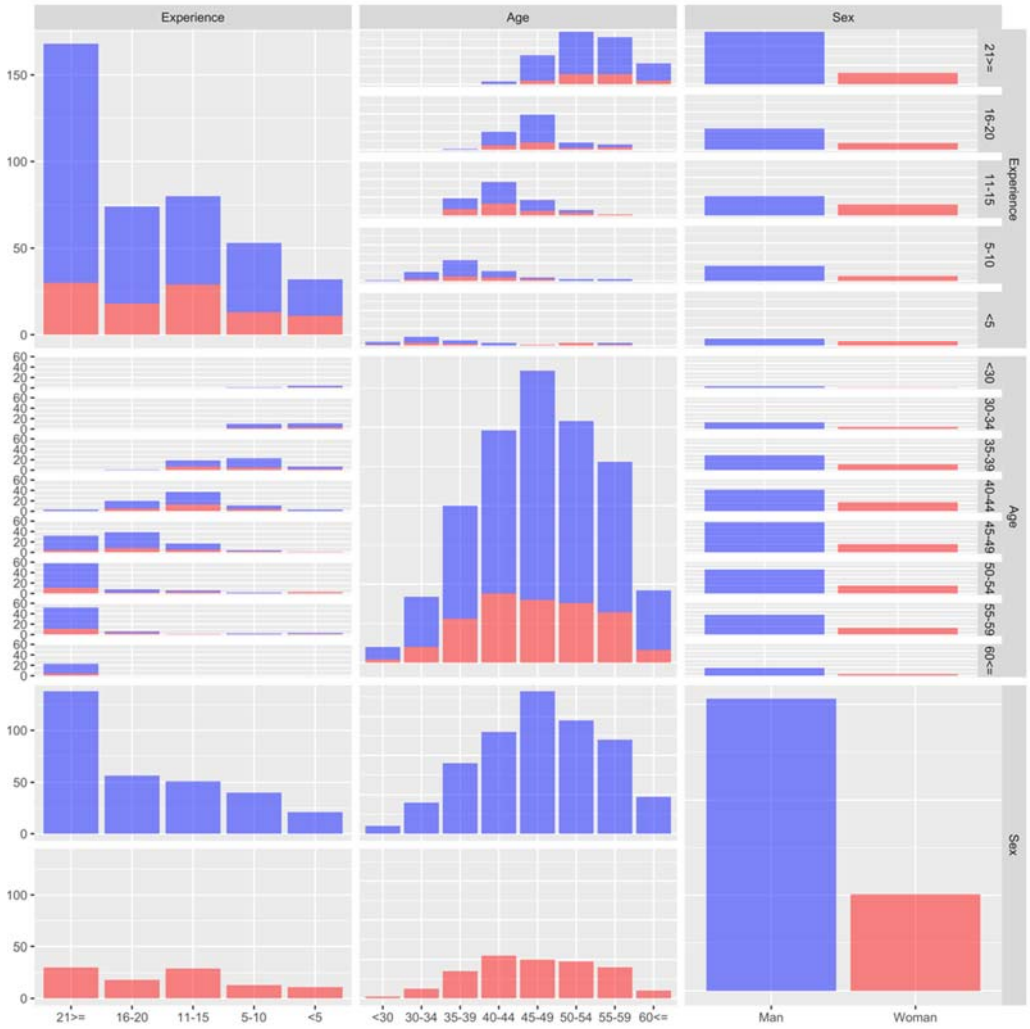


図 1 研究年数、年齢、性別の相関図

表 3 公開データを検索する時の手段 (Q F1)

ツール/情報源	人数	比率
サーチエンジン（Google など）	372	22.3%
論文や学術記事の参考文献	309	18.5%
研究者や同僚に尋ねる・教えて 貰う	186	11.1%
出版社や学術雑誌のサイト (E lsevier,Wiley な ど)	176	10.5%
学術機関のリポジトリ・データ アーカイブ	129	7.7%
データ情報のデータベース（D ata Citation I ndex など）	125	7.5%
政府・国際機関・出版社などの 広報・ニューズレター	96	5.7%
学術系 SNS (Mendel ey, Research G ate など)	79	4.7%
メーリングリスト	44	2.6%
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	43	2.6%
ブログや一般的な SNS (F acebook,Twitte r など)	38	2.3%
出版社のリポジトリ・データア ーカイブ	37	2.2%
アラートサービス (RSS 等)	25	1.5%
利用していない	11	0.7%
回答者数	回答者数	1670

表 4 経験年数による検索手段の違い (Q F1)

検索手段	経験年数 5年未満	5～10年	11～15年	16～20年	21年以上	合計
サーチエンジン（Google など）	32	44	72	67	157	372
データ情報のデータベース（D ata Citation I ndexなど）	10	15	22	13	65	125
学術機関のリポジトリ・データ アーカイブ	13	15	24	28	49	129
出版社のリポジトリ・データアーカイブ	4	4	5	4	20	37
出版社や学術雑誌のサイト(E lsevier,Wileyな ど)	13	23	37	30	73	176
論文や学術記事の参考文献	23	38	63	55	130	309
データジャーナル	4	5	9	7	18	43
政府・国際機関・出版社などの 広報・ニューズレター	11	10	20	19	36	96
ブログや一般的なSNS (F acebook,Twitte rなど)	7	7	9	5	10	38
学術系SNS (Mendel ey, Research G ateなど)	8	10	18	14	29	79
アラートサービス(RSS等)	2	4	4	4	11	25
メーリングリスト	4	6	7	8	19	44
研究者や同僚に尋ねる・教えて貰う	20	20	31	34	81	186
利用していない	0	5	1	4	1	11

最もよく使われているものは、インターネットで公開されているサーチエンジン(22.3%)、論文や学術記事の参考文献(18.5%)、研究者や同僚からの情報提供(11.1%)、出版社や学術誌のサイト(10.5%)であった。経験年数を通した検索手段の分布もほぼ同等であったが、経験年数が長い研究者はデータ情報のデータベースや研究者ネットワークを活用している様子が伺える(表 4)。しかし、オープンデータへの直截なリンクを提供するデータジャーナルは学術系 SNS やブログ等とも比較して低調であり、データジャーナルの知名度が依然として低い様子が伺える。

表 5 データの入手方法 (Q F2)

データの入手先	人数	比率
論文データベース（Pubmed Central,J-Stage, 医中誌）	276	21.6%
出版社（Elsevier、Wiley-Blackwell、Springer など）	215	16.8%
厚生労働省や総務省統計局などの政府統計	212	16.6%
学術データアーカイブ	139	10.9%
個人や研究室の Web サイト	102	8%
国際連合や OECD などの国際機関の Web サイト	95	7.4%
学術系 SNS(Mendeley,Research Gate など)	74	5.8%
特定研究機関のリポジトリ・デ ータアーカイブ	65	5.1%
コード共有サービス（GitHub 等）	45	3.5%
プレプリントサーバ	44	3.4%
国際的な研究グループの Web サイト	7	0.5%
データ共有サービス (figs hare,zenodo など)	2	0.2%
回答者数	1276	100.0%

C.2.2 公開データの入手先

公開データの入手方法を確認するため、「これまでに公開データを以下の公開先から入手して利用した経験はありますか?なお、ここでお尋ねしているのは論文ではなくデータの入手についてですので、論文誌の出版社の選択肢では、いわゆる論文の補足資料(supplementary materials)からの入手などを想定しております(Q F2)。」と尋ねた結果を表 5,6 に示す。質問では複数選択肢の他、自由記述欄を提供した。

回答は論文データベース(21.6%)を筆頭に、出版社(16.8%)、政府統計(16.6%)、学術データアーカイブ(10.9%)と続き、何らかの査読を経ているもの、あるいは法律等に基づいて信頼性のある調査方法で収集されているデータが中心的に入手されている傾向が明らかになった。学術 SNS、コード共有サービス、プレプリントサーバ等学術コミュニティで共有されているものは、低調

表 6 経験年数によるデータ入手先の違い (Q F2)

データ入手先	経験年数 5年未満	5～10年	11～15年	16～20年	21年以上	合計
論文データベース（Pubme d Central,J-St age,医中誌)	19	38	56	46	117	276
出版社（Elsevier、W iley-Blackwell 、Springerなど）	16	30	43	43	83	215
厚生労働省や総務省統計局など の政府統計	21	23	38	40	90	212
学術データアーカイブ	4	20	29	23	63	139
個人や研究室のWebサイト	12	9	24	21	36	102
国際連合やOECDなどの国際 機関のWebサイト	5	7	18	21	44	95
学術系SNS(Mendele y, Research Ga teなど)	5	9	17	15	28	74
特定研究機関のリポジトリ・データアーカイブ	5	4	12	17	27	65
コード共有サービス	2	8	13	11	11	45
プレプリントサーバ	1	9	9	8	17	44
国際的な研究グルー プのWebサイト	0	1	2	0	4	7
データ共有サービス(figs hare,zenodoなど)	1	1	0	0	0	2

に留まる。査読等を経ていないことによる品質への懸念、知名度等の問題が考えられるが、この研究の範囲では明らかではない。経験年数を通してデータ入手先の傾向は概ね共通しているが、若手は政府統計及び個人や研究室の Web サイトのデータを利用する傾向が、16〜20 年の中堅は上位３位の入手先に利用先が集中する傾向が確認された(表 6)。

C.3データ保持状況

本節では、データの保持するモチベーションや保存に関する課題、データ保持に関するスキル等に貢献すると思われる要件を確認する。

C.3.1　データを保持するモチベーション

研究終了後も研究データを保持するモチベーションについて、「貴方が生成した研究データを保持(preserve)するモチベーションについて回答下さい(公開の有無については問いません)(Q　G1)。」と尋ねた結果を図 2 に示す。質問では５段階の排他的選択肢の他、自由記述欄(Q G2)を提供した。

データを保持するモチベーションにおいて重視されている順に、将来においてデータの検証ができるようにするため(89%)、科学の発展に貢献する(88%)、既存のデータの再解析ができるようにするため(86%)、公的に支援されたものであり、公的財産であるため(84%)、他はないデータであるため(72%)、学際的な連携を推進するため(68%)であったが、経済的なメリットについては重要性を見いださない意見の方が多かった。

自由記述の回答では、以下のような記述がみられた。括弧内は分析者による分類である。

・ 迅速な意思決定を下すため、包括的なデータ情報（Big Data）は不可欠である。[再利用性、有用性]

・ 新規解析手法が実施可能になった際の検証用として重要。ただし、実験条件等の違いにより新規手法を使用することが適切ではない場合も多く、意味がある場合は少ないと考える[再検証]。

・ 当方所属機関の目的は障害者の更生・支援であり、その障害者の更生・支援に対するモチベーションは非常に重要であると考えます[有用性]。

・ 研究データは基本的に保存しておくものという考えでいるので、データを保持することに特段のモチベーションを必要としません[義務]。

・ 研究データの一定期間以上の保持を所属する研究機関で義務付けられているため（非常に重要である）論文掲載の際に研究データの公的データベースへの登録を求められるため（非常に重要である）[義務]

・科学的な検証をする程ではないにせよ、時間がたってからもう一度自分のデータを見直すという課程は非常に科学的にも心情的にも向上心をくすぐるものである。その作業によって、新たな発見（またはすでに公知の事実であったとしても、その再発見）があるかもしれない。最も重要な点は、それが公には明確な価値を持っていないとしても、発見・再発見または発見しなかった、という

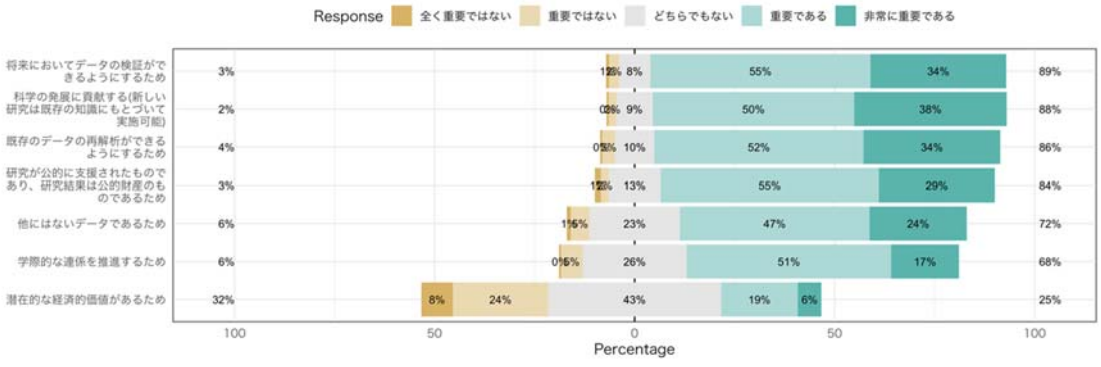


図 2 データを保持するモチベーション(Q G1)

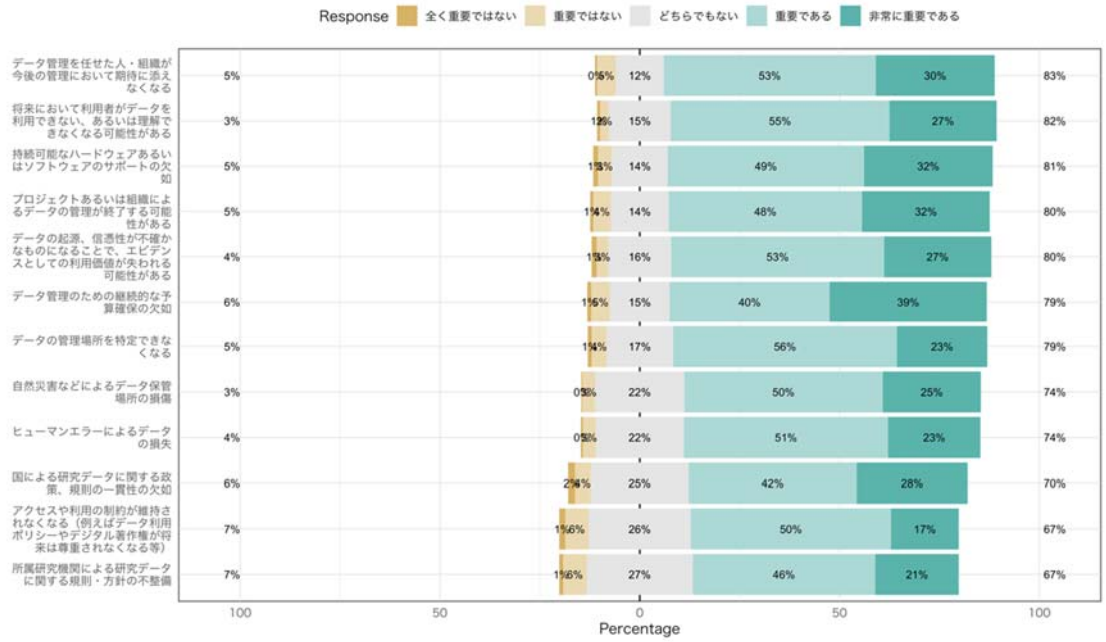


図 3 データの保持に関する課題 (Q G3)

データ見直しのプロセスそのものが、新たな知見や発想を得る手がかりとなることが多い。よって、データが個人の環境に左右されず（PC や職場が変わっても、同じようにデータを参照できる）、時間を経ても尚再利用可能であるということは極めて重要であると考える。［再利用性、再検証］

・ 研究の記録として保持しつづけること自体が重要である。［義務］

・データを持っていると、個人情報などの管理の手間があるために、早く廃棄をしたいということもあり得ます。［セキュリティ］

C.3.2 研究データの保持に関する課題

「研究データの保持に関する課題について回答して下さい(Q G3)。」と尋ねた結果を図 3 に示す。質問では 5 段階の排他的選択肢の他、自由記述欄(Q G4)を提供した。全ての選択肢において重要であるという回答が得られ、経験がないオープンデータの公開にむけて、いずれの方面にも課題を感じている研究者の姿が浮かぶ。上位 5 位において、データ管理者・組織の継続性(83%)、将来のデータ利用可能性(82%)、ハードウェアあるいはソフトウェアの持続可能性(81%)、プロジェクトの継続性(80%)、データの起源・信憑性の喪失による価値の消失(80%)と、データそのもの及びデータに関わる環境の継続性への懸念がもっとも高い。我が国では長期間に渡って安定して措置される研究予算に乏しいことは研究者であれば強く感じている方が多いと思われるが、それを裏付けるような傾向であると思われる。また安定した予算背景がないことは、管理者・組織やプロジェクトの継続性、将来のデータ利用可能性及びハードウェア、ソフトウェアの持続可能性のための維持管理、ひいてはデータの起源・信憑性の消失につながるという点において通底した課題であろう。

自由記述の回答では、以下のような記述がみられた。括弧内は分析者による分類である。

・ データの品質管理の概念が不足している．第三者がデータを利用するためのメタデータの管理がなされていないことが多い．データシェアを目的としていながら，標準の利用が少なく，データ・アーカイブにとどまっている．［再利用性］

・ フォーマット自体が古くなり、最新の科学データの登録にそぐわなくなること（重要である）［再利用性］

・ 「研究データ」といっても、データの

種類によって保管・保持の重要性・必要性は異なります。
試薬キットを用いた予備実験のデータのようなものと臨床研究データのような再取得が難しいデータとでは扱いが違います。どちらを想定しているのかがわからず、だからといって同列に回答することもできず、回答は必ずどちらか寄りにバイアスがかかります。ただしいずれにせよ、後者のようなデータの保管については、公的資金による支援の形式や期間、および持続可能性のあるビジネスモデルについてのビジョンを国が持たないといけないと思います。[事業継続性]

・ 倫理指針との兼ね合いやインフォームドコンセントの段階でどの程度、説明するのかなどについては明確なルールを規定すべきと思います。例えば、現在データセットが公開されている研究に参加した人の中で、どのくらいの人が自分のデータが web 上に保存されており、第 3 者がアクセス可能な状態になっていると知っているでしょうか？これは倫理を厳密にしてほしいといっているわけではありません。研究者個人のルールに任せて後で問題になるケース、ならないケースがでるのは避けたいので、統一ルールがほしいという意味です。例えば、インフォームドコンセントの段階で、研究発表後のデータセット公開についての詳細に触れなくても、(たとえ世界中の人がアクセス可能でも) 匿名データであれば問題なしという共通ルールがあれば対応しやすいです。一方で、現在の各機関における研究倫理委員会は、研究者に対して、COI のあるなしにかかわらず、一律の倫理基準を設けている場合や非常に詳細な倫理要綱の形式的な厳守が求める場合が散見されます。データセット公開に関して、現在の倫理手順に加えての作業があるとなると、敬遠したいということが本音です。[倫理・セキュリティ]

・ 当所では経験がないが、外部からの不正アクセスによるデータの盗難、消失など大変重要である。[セキュリティ]

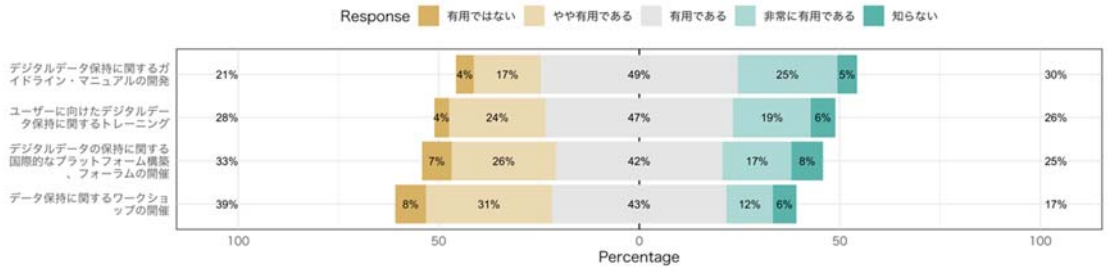
・ データの価値を素早く正確に評価することは、多面的な側面から難しい。

・ 上記：将来において利用者がデータを利用できない、あるいは理解できなくなる可能性がある（フォーマット、意味、アルゴリズム等がわかりやすい形で継承されない）に近いが、デファクトスタンダードの変更も大きな壁となっている。例えば、研究内容をテキストとして文字コードEUC-JPで保存したが、現在のスタンダードは UTF-8 であり、環境によっては参照が困難になる。次に、考えられる問題点は、新しい技術、規格に全てのデータソース管理がついて行けないという問題があると考える。個人またはアカデミアの環境を利用し、HTTP（80 番）でアクセス出来るような環境に研究データを保存し、その検索を perl をもちいた CGI で行える環境を構築した。しかしながら、当時作動していた環境から、CGI のアクセスセキュリティーポリシーが変更になり、perl のバージョンが変更になり、検索機能が損なわれてしまった。

全てのデータについて、その時代にあわせて perl、CGI 等の更新をすることができず、一部のデータ利用が不便になってしまった。[再利用性]

・ 上記の回答の多くと関連することですが、将来の検証のためのデータ保存とデータの管理(特に個人情報保護の面)とが競合することです。ここ 10 年ほどで大きく報道された研究不正が相次ぎ、データの保存について定めたガイドラインや組織内の規程が急速に整備されました。一方、私の研究領域に関連の強い医学・健康科学の分野では、研究対象者の個人情報を保護するために、それ以前からデータの管理について研究倫理指針などで厳しく定めています。研究対象者へは、研究終了 3 年後にデータを破棄する、最後の研究成果公表後は速やかにデータを破棄するなどと説明し、その内容に同意していただいています。ところが、組織の規程でデータは最低 5 年間保存するな

図 4 データ保持に関する知識増加に有用な事項 (Q G5)



どと定められていると、そちらが優先されるとの見解で、「最低」であることから実際には破棄されることがありません。そのうちデータが管理されなくなっていきますが、研究の業界では人の異動や多機関での共同研究が多いにもかかわらず、管理は容易にできるという甘い認識が跋扈しているように感じます。データの保存を重視しすぎるがために、データの管理がないがしろにされている現状を強く危惧しています。[セキュリティ]

・ データベースと称するマイ・データベースが氾濫しており集約性が欠如している。[再利用性]

C.3.3 データ保持に関する知識を増やすために有用な事項

「データ保持に関する知識を増やすために有用だと思われる事項を回答下さい (Q G5)。」と尋ねた結果を図 4 に示す。質問では５段階の排他的選択肢の他、自由記述欄(Q G6)を提供した。デジタルデータ保持に関するガイドライン・マニュアルの開発(36%)、ユーザーに向けたデジタルデータに関するトレーニング(26%)、デジタルデータの保持に関する国際的なプラットフォーム(25%)、データ保持に関するワークショップの開催(17%)の順であった。

自由記述の回答では、以下のような記述がみられた。括弧内は分析者による分類である。

・ 個人での知識増加は重要かと思うが，具体的に保持方法や，保持するための仕組みを，研究機関レベル，国レベルといった大きな枠組みでガイドライン・マニュアルを含めて整備することが重要と考えられる．企業のように組織内で共有されるデータ保持のためのサーバが整備され，さらにそのデータがインターネット技術等で共有されると良いかもしれない．環境が整備されれば，何らかのインセンティブまたは義務化が必要かと思うが，個人でのオープンサイエンス実践は自然に行われるようになると思う．[ガイドライン・インセンティブ・義務化]

・ 科研費等の研究費募集要項において、データ保持のための研究項目を用意・拡大すること（現状でも研究項目が存在するが規模が小さく限定的と思われる)。[研究領域開拓]

・トレーニング等は全て有効であると思いますが、前述の倫理やインフォムードコンセントなどの実質的な手続きや業務の煩雑さのほうがより重要です。[業務負担軽減]

・ ＩＳＯ、ＩＥＣ、ＪＩＳなどによる規格化[規格化]

・データの利活用の意義や重要性[啓蒙]

・ 研究のデータが、wet のデータと dry のデータの保持が混在しているように思われる[1]。dry のデータの保持は、より簡便に公開データベースに登録できること、また登録による研究者側の見返りがある程度認められれば、問題が解決されると思われる。

[1] Wet と Dry についての公式な定義はみあたらないが、大まかに言えば、生体から採取されるサンプルそのものを扱う、生物学的な研究(Wet)と、その試料から得られたデータを元に扱う、理論構築やコンピュータ解析を伴う研究(Dry)といった使い分けで用いられるようである。

［対象の明確化、インセンティブ］

C.4データ保持の対象

本節では研究者が生成、使用しているデータの種類や研究者が従事している研究分野について確認している。

C.4.1 使用しているデータフォーマット

「研究遂行上でよく使うデータ等のフォーマット種類について、当てはまるものについて全てチェックを入れてください(Q H1)。」と尋ねた結果を表 7 に示す。質問では複数選択肢の他、自由記述欄を提供した。

表 7 使用しているデータのフォーマット (Q H1)

データフォーマットの種類	人数	比率
オフィス文書	386	16.2%
画像	303	12.7%
汎用の表計算ソフトウェアで読 み込み可能なデータ書式	260	10.9%
ネットワーク上のデータ	188	7.9%
解析ソフトウェア専用のデータ 書式	182	7.7%
Raw データ	178	7.5%
アーカイブされたデータ	168	7.1%
テキスト	161	6.8%
構造化されたテキスト・データ	109	4.6%
設定データ	84	3.5%
マルチメディアデータ	77	3.2%
構造化されていないテキスト・ データ	76	3.2%
構造化されたグラフ	73	3.1%
医用画像 (DICOM 等)	68	2.9%
ソースコード	64	2.7%
回答者数	2377	100.0%

利用しているデータの種類は上位からオフィス文書(Word, PowerPoint 等)(16.2%)、画像(JPEG,GIF,PNG 等)(12.7%)、汎用の表計算ソフトウェアで読み込み可能なデータ書式（CSV 形式など）(10.9%)、ネットワーク上のデータ(Web サイト、電子メール、チャッ

表 8 研究者が所属する研究分野 (Q H2)

人文学		4
	人文地理学	2
	哲学	2
化学		19
	基礎化学	2
	材料化学	4
	複合化学	13
医歯薬学		470
	内科系臨床医学	126
	基礎医学	105
	境界医学	22
	外科系臨床医学	24
	歯学	7
	看護学	9
	社会医学	83
	薬学	94
工学		15
	プロセス・化学工学	2
	土木工学	3
	建築学	1
	機械工学	4
	電気電子工学	5
情報学		40
	人間情報学	12
	情報学フロンティア	14
	情報学基礎	8
	計算基盤	6
数物系科学		2
	物理学	2
環境学		10
	環境保全学	1
	環境創成学	1
	環境解析学	8
生物学		56
	人類学	1
	基礎生物学	13
	生物科学	42
社会科学		57
	心理学	20
	教育学	5
	法学	2
	社会学	25
	経済学	5
総合人文社会		4
	地域研究	4
総合理工		5
	ナノ・マイクロ科学	2
	応用物理学	2
	計算科学	1
総合生物		133
	ゲノム科学	42
	実験動物学	8
	神経科学	31
	腫瘍学	52
複合領域		83
	デザイン学	1
	人間医工学	42
	健康・スポーツ科学	6
	地理学	1
	子ども学	1
	生体分子科学	7
	生活科学	8
	社会・安全システム科学	5
	脳科学	12
農学		21
	動物生命科学	9
	境界農学	1
	農芸化学	11
総計		919

トのテキストデータ)(7.9%)、解析ソフトウェア専用のデータ書式 (SAS,SPSS 等統計処理ソフト用フォーマット等)(7.7%)であった。データの再利用性に注目して分類したFive Star Open Dataの定義[11]において、再利用可能なオープンデータとして推奨される水準に達している可能性があるものは、汎用ソフトウェアで読み込み可能なデータ形式と解析ソフトウェア専用のデータ形式、構造化されたテキストデータ（合計 23.2%）である。また画像に関する機械学習の潜在的対象となる、画像、医用画像（合計 15.6%）も確認された。

C.4.2 研究者が所属する研究分野

研究者が所属する研究分野について、国立研究開発法人科学技術振興機構の研究分野一覧の分野-分科-細目名から構成されるマスタより最大5個まで選択頂いた(Q H2)。その結果を分野-分科別に集約したものを表8に提示する。

延べ回答人数は1030人、そのうちシステムの不備による無効回答を除く有効回答は919人であった。医歯薬学(470名)、総合生物(133名)、複合領域(83名)、社会科学(57名)であった。

C.4.3 データの分野

「研究でよく使うデータの分野について、当てはまるものについて全てチェックを入れてください。(Q H3)」と尋ねた結果を表9に示す。質問では複数選択肢の他、自由記述欄を提供した。

実験（操作・介入を含む）(26.1%)、観察（非介入）(17%)、政府統計、国際機関の統計(11.3%)が過半数を占めていた。自由記述では、診療に派生して発生する情報(病院の臨床データ、手術記録、医用画像)、臨床研究(介入研究~臨床試験、治療介入データ)、研究倫理に関する書類(倫理審査書類、同意書等）の存在について指摘された。

表 9 データの分野 (Q H3)

データの分野	人数	比率
実験（操作・介入を含む）	278	26.1%
観察（非介入）	181	17%
政府統計、国際機関の統計	120	11.3%
データ・分析モデルの開発	99	9.3%
生物学的調査	95	8.9%
社会科学調査	74	6.9%
インタビュー	66	6.2%
文献資料	66	6.2%
行政記録・行政資料	62	5.8%
非生物的調査	15	1.4%
リモートセンシングされた生物データ	5	0.5%
リモートセンシングされた非生物データ	4	0.4%
回答者数	1065	100.0%

C.4.4 データの対象物

「研究データの対象物について、当てはまるものについて全てチェックを入れて下さい。(Q H4)」と尋ねた結果を表10に示す。質問では複数選択肢の他、自由記述欄を提供した。

人や動物を対象とする観察データ(12.7%)、実験データ(10.5%)、生体。生物由来物(8.3%)、ゲノム(8.1%)、細胞(7.6%)、生体情報(6.6%)で過半数を占め、続くデータも生体由来の情報が中心である。いずれも機微な個人情報が含まれる可能性があるデータである。・自由記述では、政府広報等の情報、労働災害,製品事故の発生状況のわかるもの、診療記録、臨床研究（介入研究～臨床試験）に関わるデータ、実務データが指摘された。

表 10 データの対象 (Q H4)

データの対象	人数	比率
観察データ	268	12.7%
実験データ	223	10.5%
生体・生物由来物	176	8.3%
ゲノム	172	8.1%
細胞	161	7.6%
生体情報	140	6.6%
生体反応	124	5.9%
組織・器官	123	5.8%
行動	100	4.7%
化学物質（分子）	99	4.7%
医用画像（医用波形含）	84	4%
生活習慣・ライフスタイル	80	3.8%
シミュレーションデータ	68	3.2%
意識	64	3%
社会経済情報	46	2.2%
ライフヒストリー	46	2.2%
マクロ的情報（国、地域ごとの 集計）	44	2.1%
公的支援	39	1.8%
社会経済的地位、現状	34	1.6%
収入・資産	25	1.2%
回答者数	2116	100.0%

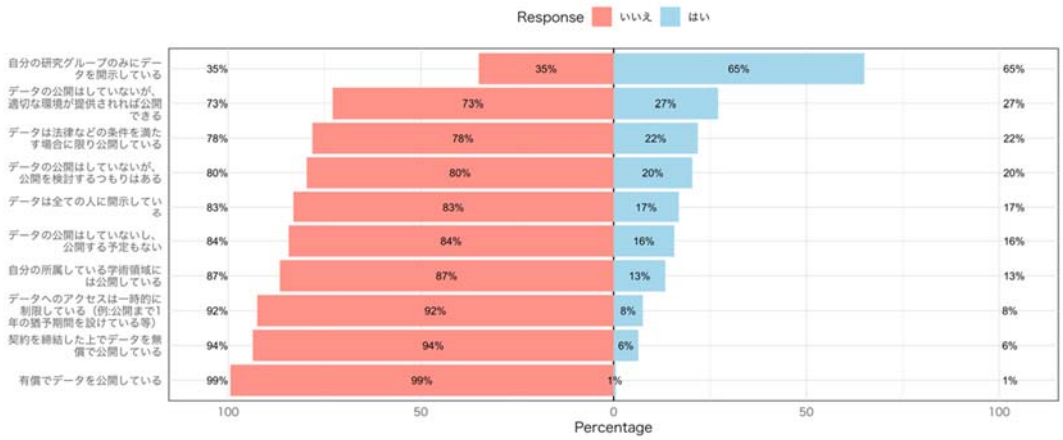


図 5 データの公開状況 (Q I1)

C.5学際的な研究データの利用

C.5.1 データの公開状況

「データの公開状況について選択してください。(複数選択可能)保有するデータが複数あり、公開に関する状況も異なることが考えられますので、当てはまるものを全て選んで下さい）（Q I1)」と尋ねた結果を図 5 に示す。質問では複数選択肢の他、自由記述欄を提供した。自分の研究グループのみにデータを開示している（65%）は研究遂行上必須であるので当然の結果であるが、それでも 35%は研究グループでも公開していないことは、機微な情報を含むデータがあるが故に可及的にデータを保管する場所を増やさないようにしている可能性も考えられる。適切な環境が提供されれば公開する(22%)、現在公開はしていないが公開を検討する積もりはある(20%)という意見があるものの、全体的には公開に対して非消極的な姿勢が伺える。一方で、データの公開はしていないし、公開する予定もない(16%)が少ないため、必ずしも将来においても公開に対して否定するものではないというスタンスも垣間見える。

以下に自由記述欄の回答を示す。

・ 一部データを法人ホームページに掲載

・ 公開されているデータを用いた研究を行っているため，いずれにも該当しない.

・ 生データは公開していないが発表 PPT のようには公開している場合がある

・ 公開しているデータもあるし、公開していないデータもある

・ 有用なデータを多数公開しているが、公開していないデータもある（複数選択可とあるが、一部項目を選ぶと他がグレーアウトするため、その他に記載した）

・ データの公開は厚労省の審査が通ったものだけを公開請求者に開示予定である

・ 所属機関のデータポリシー策定とリポジトリ整備が完了したら対象となるデータを公開する。

・ 論文掲載によって公開

・ 論文化にあたり公開を求められた場合に全てのヒトに開示する形で対応した

・ 学術誌への投稿

・ 論文を発表した学術誌の規定に基づき請求に応じて個別に公開する

・ データは公開していないが、一部のデータなら公開を検討するつもりはある

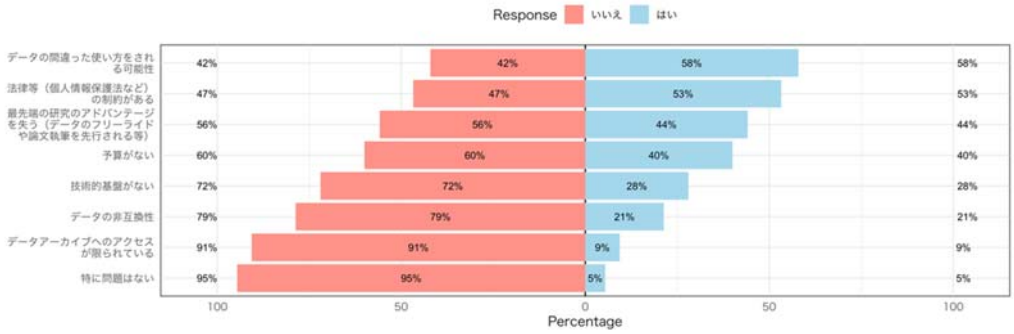


図 6 データ公開の障壁 (Q I2)

C.5.2 データの公開に関する障壁

「今後、データの公開に関する障壁となる可能性があるものについて記載して下さい(複数回答可) (Q I2)。」と尋ねた結果を図 6 に示す。質問では複数選択肢の他、自由記述欄を提供した。過半数を超えた意見として、データの間違った使いかたをされる可能性への懸念(58%)、法律上の制約(53%)が確認された。

自由記述欄の回答を以下に示す。

・ データの真正性が担保されないデータの混入による信頼性の低下。忍耐が必要な研究が軽視され、解析のみや発信力が高い人間への研究費の集中などが生じうること。

・ 共同研究者間で考え方がデータ公開への積極性が異なる．全員の許可を得る必要がある場合には普段の業務の中で時間を割くことが難しいと思う．

・ データを承諾なしにメタアナリシスされる可能性

・ 地方自治体との協定の中でいただいたデータなので、研究者以前に自治体の公開に対する考え方が反映される

・ データ収集が最も労力のかかるところであり公開には厳密な契約が必要である。

・ 倫理指針､インフォームドコンセント、参加者の真の理解、研究者の責任の兼ね合いと明確なルールの規定、および業務内容の煩雑化の防止

・ 質問の意味がよくわかりません。論文のデータは公開されますが、それとは別の話でしょうか。ルールとして､プレプリント以外に公開してしまったデータは論文に使用できないと思います。

・ 特許性を失う

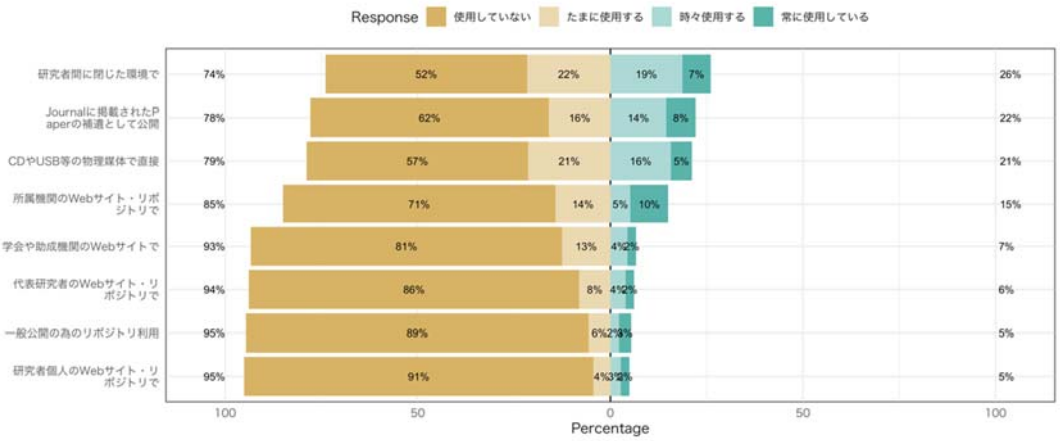
・ 国内外で公開によるフリーライドが頻繁に発生するため。

・ 面倒

・ データ公開を維持する手間と人的リソースの欠如

・ 公開することを前提として調査データを収集していない

図 7 研究データの提供状況 (Q J1)



・ 特許等の利権が絡む

・ 公開をしてもデータ収集の労に報われない。二次活用者の態度が悪い

・ データの取得方法、厳密に見えるデータ定義の曖昧さ等、データ収集に係わる現実の複雑性を理解するための研究経験が少ない研究者がデータを誤用すること

C.6研究データの提供状況

C.6.1 研究データの提供状況

「データを提供している場合は、どのような形態を使用していますか。(Q J1)」と尋ねた結果を図 7 に示す。質問では排他的複数選択肢を提供した。研究者間に閉じた環境(26%)、Journal に掲載された論文の補遺として(225)、CD や USB 等の物理媒体(15%)、所属機関の Web サイト、リポジトリ(15%)等が主流であり、外部機関のリポジトリの利用経験は低調であることが確認された。また以下の自由記述欄の回答に見られるように、学術分野によっては、その分野において確立されたリポジトリやデータベースがある場合は、そこに寄託することがあることが確認された。

「データの提供形態が上記に当てはまらない場合、公開方法について 具体的に記述下さい。(Q J2)」と尋ねた結果を以下に示す。

・ 直接請求者に開示せず、厚労省の策定する方法で開示予定です。

・ NBDC の database archive に寄託。

・ EGA、NBDC

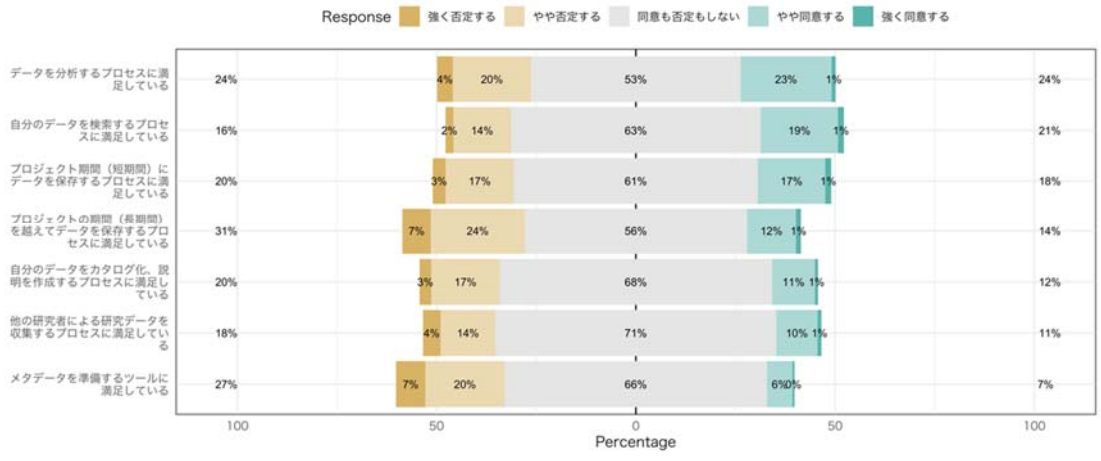
C.6.2 データ収集・準備プロセスへの満足度

「貴方が研究データを収集および準備・使用する方法についてお尋ねします。(Q J3)」と尋ねた結果を図 8 に示す。質問では排他的複数選択肢を提供した。全体的にプロセスの満足度に可も不可もない回答が中心であるが、満足が不満を超えている分野がない。なかでも「プロジェクトの機関を越えてデータを保存」、「メタデータを準備するツール」、「データをカタログ化、説明するプロセス」の順に不満が満足を上回っており、この分野でのサポートが望まれていることが伺える。

C.6.3 組織・プロジェクトのデータへの関与

「貴方の所属している組織・プロジェクトがデータについてどのように関与すべきと考えているかについてお尋ねします。(Q J4)」と尋ねた結果を図 9 に示す。質問では排他的複数選択肢を提供した。先の設問に関する状況を裏付ける形で、組織・プロジェクトが関与すべき課題は全方位に渡って存在していることが確認された。

図 8 データ収集・準備プロセスへの満足度(Q J3)



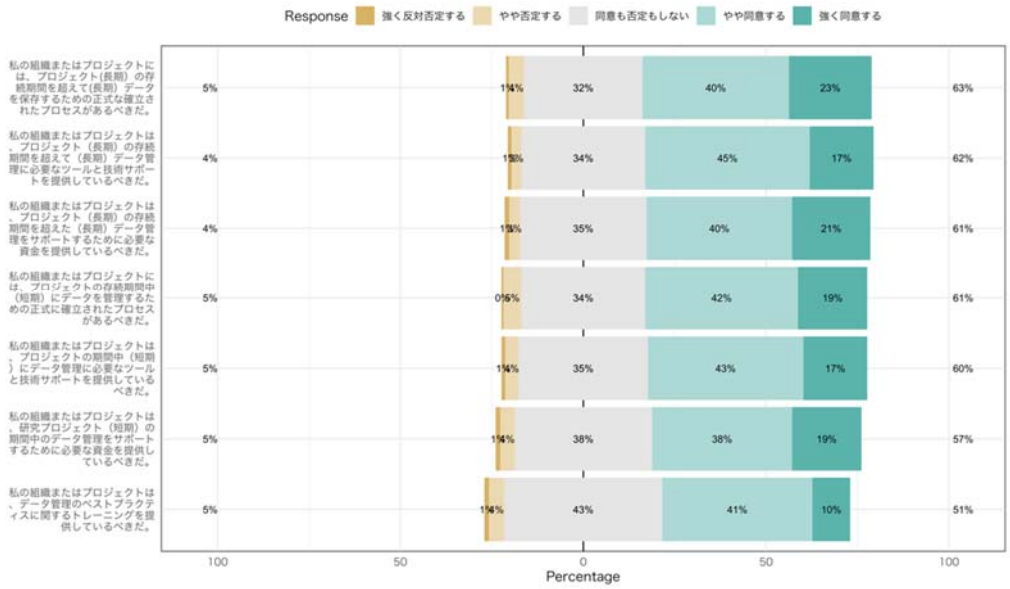


図 9 組織・プロジェクトのデータへの関与(Q J4)

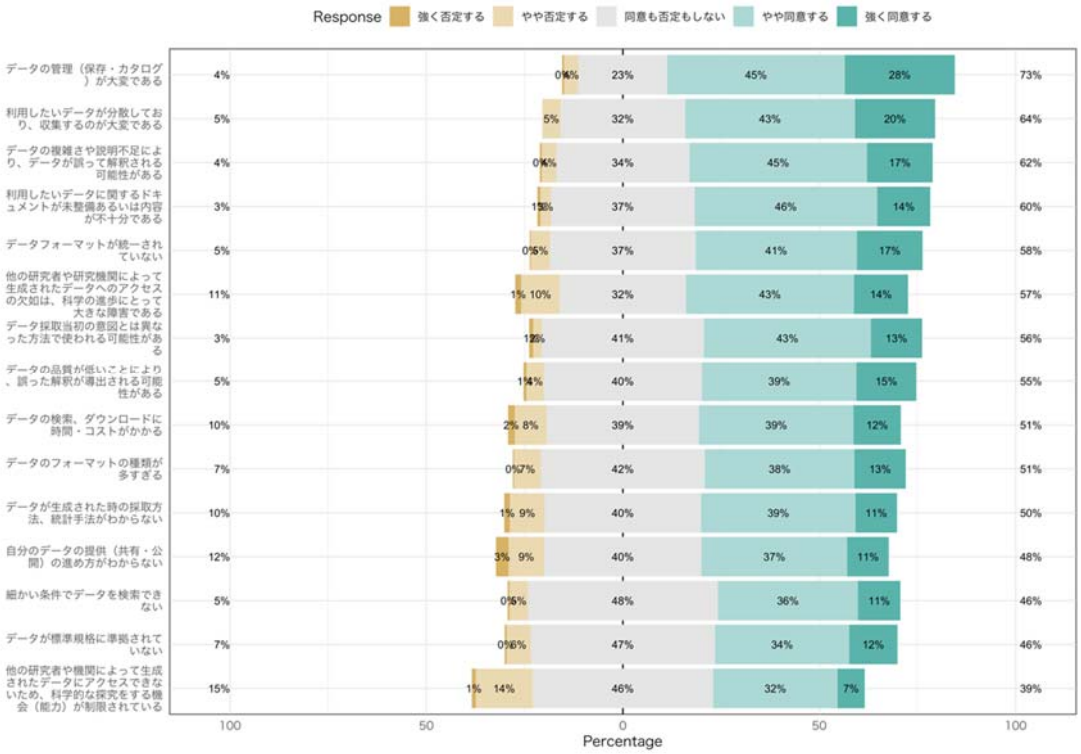


図 10 データ利用に関する課題・懸念(Q J5)

C.6.4 データ利用に関する課題・懸念事項

データの管理の負荷(73%)、データの散在(64%)、データの誤った解釈(62%)、データの説明の不備(60%)、データ標準形式の不在(58%)が上位の懸念として確認された。また、それに続く過半数を超える意見も踏まえると、「データの誤った解釈」「データの標準形式への準拠不足」の 2 つの大きな課題として集約されると思われる。

「貴方が所属している研究分野でのデータ利用に関しての課題、懸念事項について意見をお伺いします。(Q J5)」と尋ねた結果を図 10 に示す。質問では排他的複数選択肢を提供した。また、「貴方が所属している研究分野でのデータ利用に関しての課題、懸念 事項について具体的にご記載くださ

い。(Q J6)」と尋ねた結果の、自由記述欄の回答を以下に示す。

・データの利活用されることを考慮したデータ収集がされることが課題である.

・質の悪い研究で誤った結論が提示される。

・生体や人体から取得したデータ、特に脳活動の生データに関しては、解析法の進展により将来的に個人の特定やその情報を大量に抽出可能となることが否定できず、基本的に共同研究者以外には公開するべきではないと考える。共同研究者の範疇を越えて提供する場合はそれなりに厳格な契約が必要ではないか。論文においてそれほど影響力がない研究グループが開発したものを影響力が大きなグループが引用し、その引用した側の論文が大量に引用され、もとの論文を書いた研究グループが十分な評価が得られないこともある。データの過度な共有は同様の例を増やす可能性もあり、短期的な発展ではなく長期的な研究の多様性維持を重要視することも必要ではないか。

・データを効率良く保存管理し，共有・公開するためのツールが必要であると感じている.自動化することは現状難しく,整理するだけで多大な労力を要する.

・基本的に予算が不足しており、大抵のデータ提供は当該研究の研究期間が終了したり、担当者が離職してしまうとアクセスできなくなってしまう。

・データ採取当初の意図とは異なった方法で使われることは、異なる視点からの解析等により本人が見出せなかった事実を発見する上で重要なため、否定はしない。一方、特定の結論を導くために恣意的に利用される(誤った解釈を恣意的に導く)可能性も否定できないことから、その可能性を防ぐためのルール作りは必須である。

・論文に記載されているデータにもかかわらず、登録データに欠損があり、解析が困難なケースがある。データの品質の保証が課題と考える。

・データの形式が様々で、ビッグデータとして取り扱いが難しい。データクリーニングの手間がかかりすぎる。電子カルテベンダーで持ち方が様々でビックデータとしての利用を考えたつくりにはなっていない。SS-MIX２の取り組みも十分生かされていないようにみえる。

・１）この分野では、特定人のみに配布した機密情報を勝手に特許明細書として作成、公開し、特許出願をしようとする人が多数認められる（企業が特許出願にノルマを与えているため)。

２）この分野では、他の人が発明、発見した新規性のある内容(実験結果など)を勝手に手に入れて、発明者や発見者の名前を意図的に入れずに自分の成果として学術論文に発表する人がいる（某国立研究開発機関の人にそのようなことをやられたことがあった)。

・私が所属している生物系の分野では，たとえ論文になっている結果であっても，再現性の低いのが大問題になっています。自分の主張に合うデータも，そうでないデータも同じくらいに見つかるのが現状です。データ公開は基本的には良いとは思うのですが，私の属する分野でもしそれを促進した場合,研究が進むことよりも,データの質が玉石混淆となり，質の悪いデータを用いた研究者が謝った結果を導く可能性の方を強く懸念します。当該分野では,ある種の「流行」が出来上がると，それに追随するようなデータ,論文が増えます。それは時として，雑に実験を行えば得られる結果であったりします。ですから,非常に丁寧に実験をして得た,精確性の高いデータが,精確性の低いデータの中に埋没する可能性を懸念します。実際,或る論文のサプリメントにあるデータを信じて研究を始めたものの，再現性がなくて，論文の筆者に問いあわせたら「論文を出した後にもう一回やったら，その結果は違った」と言われたことがあります。また,自身のデータを利用されることを考えても，謝った解釈で利用されることは困ります。私のいる分野に関しては,デー

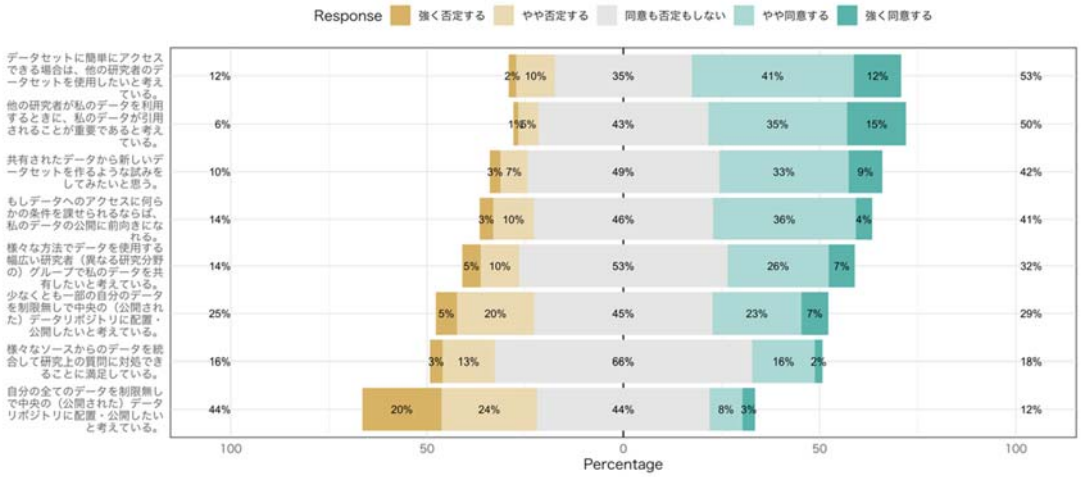


図 11 データ共有に関する意向 (Q J7)

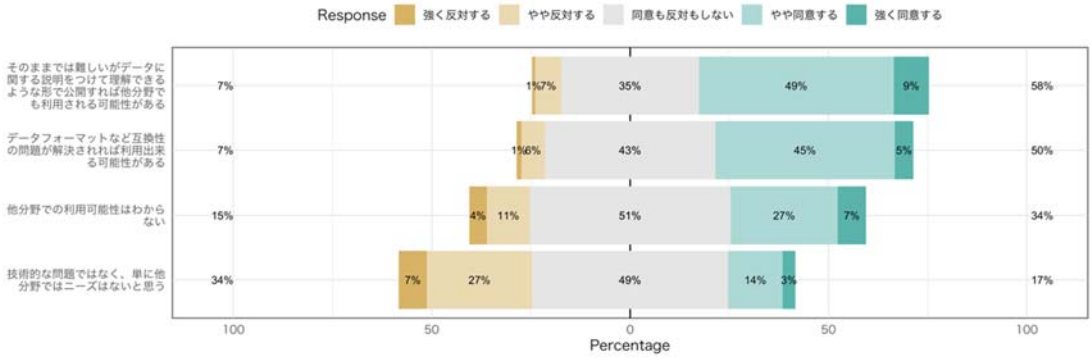


図 12 異分野の研究における利用可能性 (Q J8)

タそのものというより，何かしらの文脈がないと意味を持たないデータが多いので，公開に関しては慎重にならざるを得ません。

・公衆衛生分野におけるデータの共有や利活用の重要性および意識が低いと感じる.

・データを収集したことのない研究者が急増していることや、「真実の探求」よりも「論文が書ければよい」といった風潮が蔓延しているため、データの限界などを理解せずに P 値の解析だけをして論文を書こうとする研究者が多い。また論文のレビューアーの知識レベルが追いついておらず、全否定・全肯定など極端な態度に走りがちのレビューが多く、きちんと研究課題に照らしてデータの限界がどう影響するのかを冷静に評価できていない。また、データの限界を検証する論文など、みんなが限界を認識するので Bad News を伝えるようで嫌われる。データの共有の無さよりも、これらのインセンティブやシステムの欠如が、科学としての社会医学を毀損している。

また、データの公開を論ずる際に、データを収集しているひとへの敬意や、正当な報酬という議論が無い。データを集めること自体が手間がかかる割に、人に利用されて損をするという構造的問題がある。また、使わせたデータを間違った使い方解釈をされたら困る、では、チェックをすればよいかというと、いちいちチェックをする余裕もない。データの内容について、知識や限界の試験をして、それに通った者だけが使える、という事前チェックならまし・・。敬意の問題は残るが。

C.6.5 データ共有に関する意向

「データ共有についてお尋ねいたします。(Q J7)」と尋ねた結果を図 11 に示す。質問では排他的複数選択肢を提供した。過半数を超えた意見として、データセットに簡単にアクセスできる場合は、他の研究者の

データセットを利用したい(53%)、他の研究者が利用する時は引用されることが重要(50%)、とデータの相互利用と公正な引用による相互互恵的な運用の希望が伺えた。

C.6.6　異分野の研究における利用可能性

「データを公開した場合、異分野の研究によって利用できると思われますか?(Q J8)」と尋ねた結果を図 12 に示す。過半数を超えた意見として、「そのままの利用は難しいがデータに関する説明を付けて公開すれば利用される可能性がある。(58%)」、「データフォーマット等互換性の問題が解決されたら可能性がある。(50%)」が確認された。

質問では排他的複数選択肢を提供した。また、「データを公開した場合、異分野の研究によって利用できるか、上記 以外の回答がある場合は反対・同意も含めて記載をお願いします。(Q J9)」と尋ねた結果を以下に示す。

・他分野での利用可能性を考慮して研究することが大切になってくると考えています.

・必要十分な情報が提供されるかが重要で、そうであれば利用は可能と考える。基本的には提供情報以上の質問は一切なし、ないし代わりに質問に回答する人的リソースがあるのであればデータの範囲や内容によっては提供に同意する。登録されているものが有用なデータであればあるほど、利用者が増え、質問や対応に追われることでさらに有用なデータを生みだせる人的リソースが(ほとんどの場合)無為に消費されることが最大の問題と考える。

・ とくにありません

・ 利用可能かどうかよくわかりません。

・「共有されたデータから新しいデータセットを作るような試みをしてみたいと思う」、例えば同じ薬剤の効果を測る RCT で

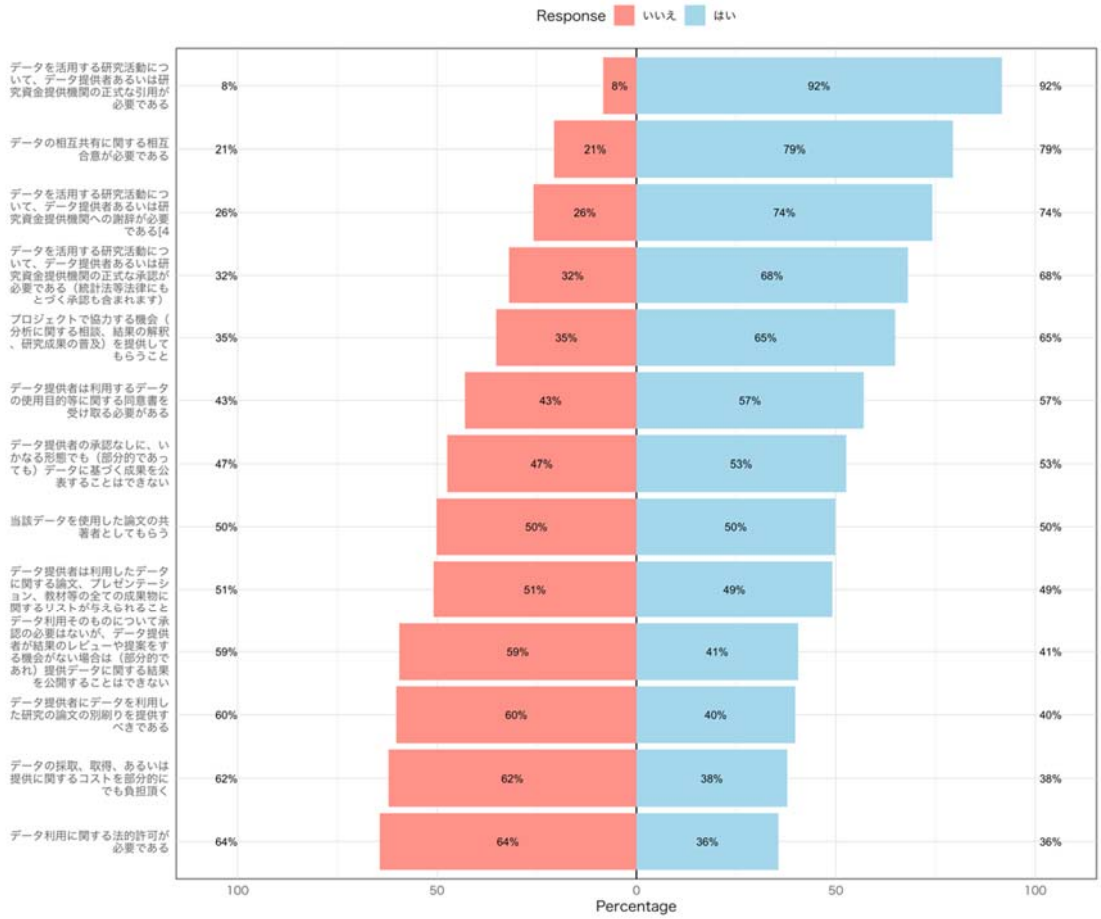


図 13 データ使用時の公正な条件(Q J10)

あっても、導入基準などが異なる場合が多いです。研究間の異質性とその対処などについて理解している者以外が作成するデータセットや分析、生み出された知見が、科学に対してむしろ悪影響を与えそうで怖いです。

・ 医学研究のデータであれば，治療や検査法など異分野の研究に利用できる可能性が高いと考えている.

・ データの品質、生体試料のサンプリング手法、治療応答の評価など、おそらく個々の研究で相当異なる。ただまとめて集めるだけで成果を上げられる研究はかなり限定的だろうと思われる。

・ 間違った使い方が増えると思います。

C.6.7 データ使用時の公正な条件

「他人が貴方(データ提供者)のデータを使用する場合を使用するに際して、公正な条件はどのようなものであるべきかのお考えについてお尋ねいたします。(Q J10)」と尋ねた結果を図 13 に示す。質問では排他的複数選択肢を提供した。

過半数を超えた意見として、データを活用した研究者による正式な引用が必要(92%)、データの相互共有に関する相互合意(79%)、データ提供者への謝辞(74%)、データ提供者による承認(68%)、プロジェクトで協力する機会の提供(65%)、データの使用目的に関する同意(57%)、公表前の承認(53%)、論文の共著者(50%)が確認された。

C.6.8 データ公開のライセンス

「データを公開する際のライセンス形態で相応しいと思われるものについて選択して下さい。(Q J11)」と尋ねた結果を図 14 に示す。質問では複数選択肢及び自由記述欄を提供した。なお、選択肢として列挙したライセンス条件は排他的ではなく、組み合わせることが可能である。「CC-継承(38%)以外は過半数の支持を受けていた。

以下に自由記述欄の回答を示す。

・ 質問の意味が分からない(同内容 3 名)

・ データ毎に適切なライセンス形態があり、特定のライセンス形態を一律で強要すべきでない。

・ 一つ前の質問もそうですが、データの種類によって回答は変わるので、きちんと回答できません。

C.6.9 データ公開のライセンス選択の理由（自由記述）

「差し支えなければ、上記のライセンスを選択する理由について端的に記述下さい。(Q J12)」と尋ねた結果を以下に示す。

・ モラルに依拠するだけではなく、コンプライアンスの意識が必要である

・ 営利目的であっても役に立つ研究もあると思うので、何らかの条件をつけることで利用可能とすべきと考える。

・ 公開するデータの内容により異なるため.

・ 犯罪･違反時については論を待たない。改変ありの場合、トラブルの基盤になりうる。

・ データアクセスの自由度と、データの意図しない利用に対するトレードオフを考えた結果。

・ CC-表示：データの出所は追跡できるようにすべき、CC-継承＆非営利はデータ毎に条件が異なる、CC-改変禁止は原則。利用形態によっては正当な改変は許可されるべきだがケースバイケース、犯罪利用は禁止

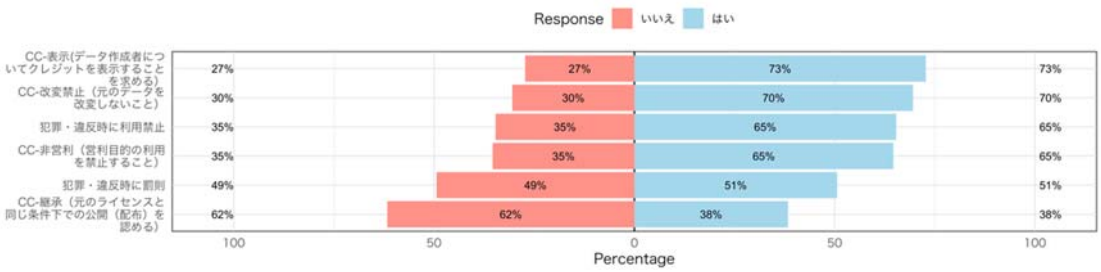


図 14 データ公開のライセンス (Q J11)

すべき、罰則は行政が行うことで、データ提供者の管轄外

・ 責任の所在を明らかにするため

・ データの恣意的な使用を防ぐ必要があること、元データに基づいた科学的且つ論理的な議論が可能となることを考慮すれば、上記の条件が必須と考える。

・ ケースによると思うが、できるだけオープンに利用できることが望ましいと考えるため。

・ ライセンスに違反したユーザーを訴える人的・経済的基盤が無いにも関わらず制限をかけるのは意味が無いから。

・ データの誤った利用禁止。あくまで科学の進展のためのソースとして公開が望ましい

・ 細かいことはよくわかりません。（同内容 1 名）

・ データ提供者への敬意と、問題が起きた時にデータ提供者を守るため

・ データの出所の透明性やデータ自体の正確性を保つために表示・改変禁止は必要と考えた。継承に関してはデータ利用者が追加のデータを加えた成果物を出したケースを考えると強すぎる制限と感じたので外した。

・ 透明性、公平性、質の担保

・ 公表されたデータの解析結果を理解するためにソースとして用いられたデータの出所とデータの状況を把握する必要がある。（バイアスなどが理解できる）

・ データはオープンであるべきだが、意図しない利用を防ぐ手立ても必要。

・ 利用の自由度をなるべく高めたい一方で、利用者が提供者になるときの公開条件は原作者（＝最上流の提供者）の意思が下流まで及ぶ必要があるため。

・ 自身のデータから他者が新しい成果を生み出したことも自身の実績として利用するため

・ 制限は少ない方が良いが、データ作成者にも何らかのメリットがあった方がよい。

・ データ提供者と使用者の立場を明確にする

・ データの目的外使用を防止し，研究・教育目的のための利用に制限するため。

C.6.10 データ利用に関するメトリクス

「データを公開した場合、関心がある項目を選択して下さい(複数選択可)。(Q J13)」と尋ねた結果を図 15 に示す。質問では複数選択肢の他自由記述欄を提供した。過半数を超えた意見は、引用数(85%)、閲覧数(71%)、ダウンロード数(65%)であった。

以下に自由記述欄の回答を示す。

・公開予定なしの為該当しない（同等内容 6 名）

・関心なし（同等内容 2 名）

・ 正しい解釈がなされているか。

・ データを元にした実際の問い合わせの有無

・ データの品質管理方法

・ 再現性という本来の目的から逸脱しますので、数値評価は絶対に避けるべきと考えます（そういった評価は論文だけで十分です)。

C.6.11 データ提供できない理由

「貴方が第三者にデータを提供しない、あるいは提供できないことがある場合につ

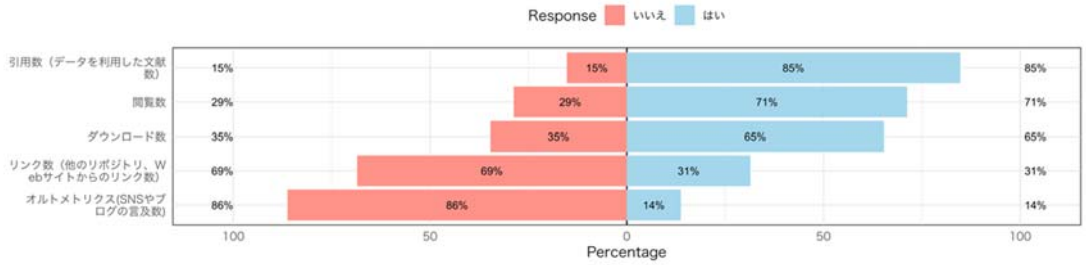


図 15 データ利用に関するメトリクス (Q J13)

いて、その理由について当てはまるものを全て選んでください。（Q J14）と尋ねた結果を図 16 に示す。質問では複数選択肢の他自由記述欄を提供した。

過半数を超えた意見は、データ公開にあたり要求される作業に対応できない（58%）、データ公開の準備をする時間がない（51%）と労力不足に伴う意見であった。以下、自由記述欄の回答を示す。

・ 個人情報保護

・ 一部のデータは（公的資金をもとにしたものに関しては）研究報告書と言う形で公開されている。また、一部のデータは所属機関のＨＰで掲載している。上記はその他、公開していないデータに関しての回答です

・ 厚労省等からの委託事業については、勝手に公開できない

・ お見せするほどのものでない

・ 公平性が整備されていないと感じる

・ 有用なデータは公開しているが、全てのデータを提供しているわけではない

・ 提供先から制御できない形で第 3 者にデータがわたることを防げない

・ 研究データとして提供してくれた自治体に承認が必要なため

・ データ提供により現在の自分の研究が他の施設に先んじられ、低い評価を受ける懸念がある

・ データを共用している共同研究者がデータを提供したがらない

・ 将来何らかの問題に巻き込まれる恐れがぬぐえない

・ データの利用者の態度が悪いと、収集にかかわったスタッフの不満がたかまり納得させられない

・ 本調査の趣旨に沿ったデータを持っていない。

C.6.12 データ提供が可能になった場合の対応

「前項の質問に関連して、「上記の理由が解決された場合、研究データを公開したい

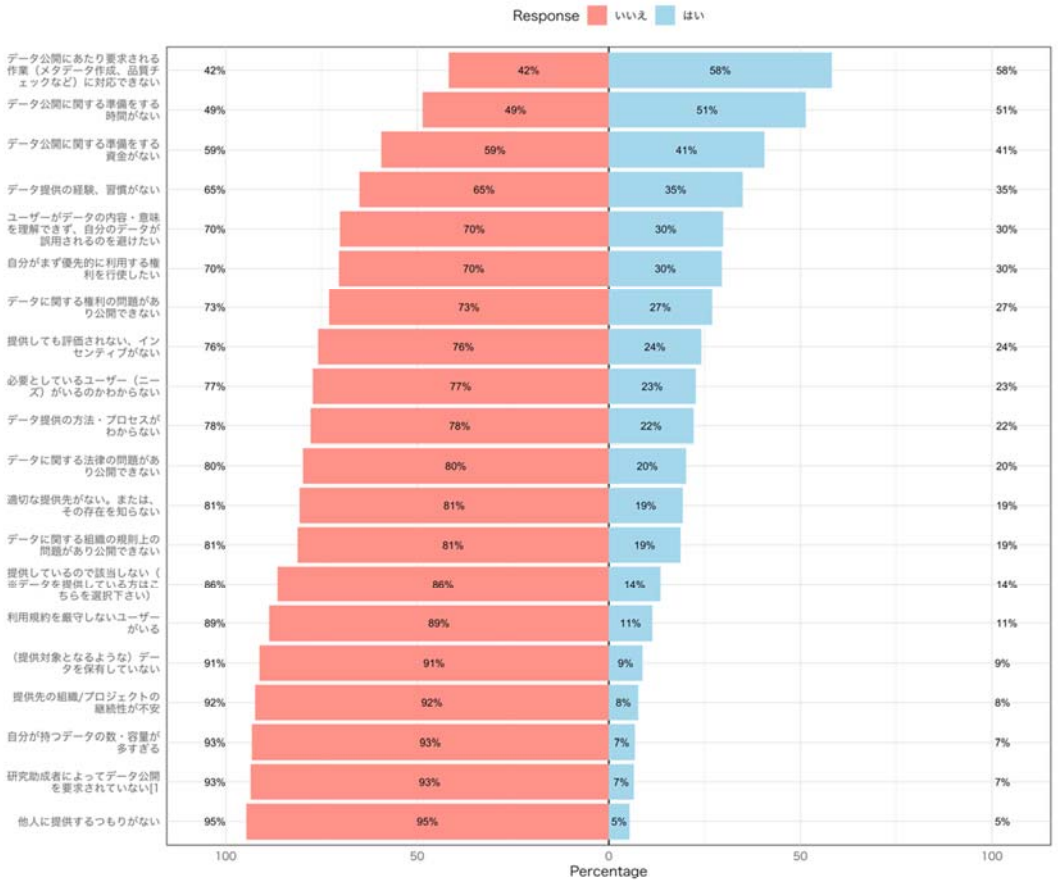


図 16 データ提供できない理由 (Q J14)

と思われますか?(Q J15) と尋ねた結果を表 11 に示す。

表 11 データ公開の意向 (Q J15)

公開の意向	人数	比率
はい	144	41.1%
いいえ	29	8.3%
わからない	177	50.6%
回答者数	350	100.0%

C.6.13 DMP の提出

「研究機関の予算以外で、主に利用している研究助成機関から、データマネジメントプラン (DMP :Data Management Plan)の提出を求められたことはありますか?適用できるものに全てチェックしてください。(Q J16)」と尋ねた結果を表 12 に示す。質問では複数選択肢の他自由記述欄を提供した。以下に自由記述欄の回答を示す。

- 学術雑誌への投稿時
- Data Management Platform？
- 書面による DMP の提出を求められたことはないが､研究成果評価委員等から DMP 類似のコメント・問い合わせを受けたことはある
- 労災疾病臨床研究補助金
- 厚労省受託事業費
- わからない
- 特定臨床研究において
- 海外の研究助成機関
- ない

表 12 DMP 提出を要求した助成機関 (Q J16)

助成機関	人数	比率
DMP の提出を求められたことはない	320	77.1%
日本医療研究開発機構（AMED）	62	14.9%
厚生労働科学研究費	14	3.4%
日本学術振興会学術研究助成基金助成金	14	3.4%
民間助成団体	3	0.7%
国立研究開発法人科学技術振興機構 (JST)	2	0.5%
回答者数	415	100.0%

C.6.14 DMP を求められた助成（自由記述）

「上記で当てはまるものがある場合、その研究課題名(助成組織含めて)についてご教示ください。可能でしたら、当該研究課題に関する説明がある Web サイトへのリンクを御提示下さい。(Q J17) と尋ねた結果について以下に示す。

研究課題名の提示は回答者の個人を特定しうる可能性があるために、本稿では提示しない。前節の質問の選択肢に収録されていない助成機関はなく、AMED に関連した研究が全てであった（10 件)。

C.7データに関する技能

C.7.1 受けたいトレーニング

「データ公開にあたり、今後受けたいトレーニング等について当てはまるものを全て選択して下さい。(Q K1)」と尋ねた結果を表 13 に示す。質問では複数選択肢の他自由記述欄を提供した。自由記述欄の回答を以下に示す。

- メタデータの管理法
- 色々と受けたいのは山々だが時間が無い。
- 包括的な知識
- ISO のような国際規格認定制度を活用する。

表 13 受けたいトレーニング (Q K1)

受けたいトレーニング	人数	比率
データの安全な管理方法（セキュリティ）	275	16.2%
適切なデータ形式	233	13.7%
知的財産権やライセンス	224	13.2%
匿名加工の基準と実施方法	198	11.7%
適切なリポジトリ	193	11.4%
データのバックアップ方法	191	11.3%
適切なメタデータ付与方法	188	11.1%
データのバージョン管理法	155	9.1%
特にトレーニングの必要はない	40	2.4%
回答者数	1697	100.0%

C.7.2 第三者による支援が必要な事項

「データ公開にあたり、御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※などの第三者による支援が必要と思われる項目 で当てはまるものを全て選択して下さい。(Q K2)」と尋ねた結果を表 14 に示す。質問では複数選択肢の他自由記述欄を提供した。以下に自由記述欄の回答を示す。

- 情報検索方法
- メタデータ管理, 収集データから一貫したデータ管理

・ わからない（同等回答 3 名）
・ データ利用者への対応。
・ 第 3 者による試料・情報の精度管理
・ ケースが想定されない

表 14 第三者の支援が必要な事項 (Q K2)

必要な支援	人数	比率
適切なデータ形式への変換	266	14.5%
知的財産権やライセンス	243	13.2%
適切なリポジトリの選択	234	12.7%
機関のリポジトリによるデータ公開手続き	231	12.6%
メタデータの作成	195	10.6%
匿名加工の実施	184	10%
適切なメタデータ標準の選択	183	9.9%
再利用性があるデータに整える	175	9.5%
データを異分野の研究者に紹介する	129	7%
回答者数	1840	100.0%

表 15 データに関する評価指標 (Q K3)

評価指標	人数	比率
データの引用回数	355	35.9%
データ提供者としてのクレジット	230	23.3%
データのダウンロード数	229	23.2%
データの共有件数（リポジトリへの登録回数）	106	10.7%
データ採取の従事時間・従事プロジェクト数	69	7%
回答者数	989	100.0%

C.7.3 データに関係した研究者の評価

「データ作成・公開に関わった研究者の業績を評価するために、これまでの論文数・IF に加えて評価すべき項目について当てはまるものを全て選択して下さい。(Q K3)」と尋ねた結果を表 15 に示す。質問では数選択肢の他、自由記述欄を提供した。自由記述欄の回答を以下に示す。

・ 分野横断的な数値で研究者を評価する考えはやめた方がいい。野球選手とサッカー選手の能力を得点数で比較するようなもの。論文の本文をしっかり読んで、プレゼンを聞き、実際に会話もして総合的に判断するというプロセスを怠るな。
・ 特許数
・ わからない（同等回答 2 名）
・ 回答不能
・ すべてを業績評価につなげる必要はない
・ 利用者との納得可能な交換条件だと思います

C.7.4 自由記述

「オープンデータ等について、意見、質問等がありましたら、ご記載ください。(Q K4)」と尋ねた結果を以下に示す。

・研究領域や公開する情報によって温度差があるかと感じています．国際的な問題や国際的な解決を行うには，オープンデータやオープンサイエンスの取り組みが必要と考えています．
・そぐわない分野もあるので適切な取捨が必要
・色々前提が違うようなので答えるのが難しいアンケートでした。
・本件について自分は，まだピンときていないところがある．
・学術論文が有料である場合が多く、これは知的財産へのアクセス制限と感じることがある。
・学術論文の参考文献リストが、各学術誌によって異なる。この点を全て同じフォーマット（掲載年、ページ番号、巻号、著者名の書き方）にすれば、新たなオープンデータとなるのではないか。
・オープンデータ概念の導入は基本的に良い方向性だと思うが、オープン化は強制されるべきではない。有用なデータの公開を評価し、そのデータの著作権の保護や公開活動への予算助成や人的支援の強化、といった施策により、自然に推進されることが望ましい。
・データ公開の意義は理解できますが研究者が常識的であることが前提です。まず、この常識性を獲得するところから始めるべきです。善意なく、労力を割かずにあわよくばデータをもらおう、研究費を出せという研究者が多いです。一部の善意あるコツコツ努力している研究者との出会いが何より貴重です。

一般的にデータは収集に最も時間と労力、リスクがかかります。ともすればデータ解析技術のトレーニングさえすれば研究者として評価されるようになりかねません。異分野でコラボレーションする機会があれば積極的に公開する必要はないと思いますし、私は現にコラボレーションしています。な

お、データ収集をしていた研究者が別の機関に異動し、そちらで研究を続けるためにデータを移管させることは賛成ですし、全く支障がないと感じます。

・評価に関しては細心の注意を払っていただきたいです。

オープンデータ等は、再現性やアクセス性などの問題を解決するために重要であると思います。そうなると、誰でもアクセスできる状況を作ることが目的であり、多くの人が引用した、ダウンロード数などは問題の本質から大きく逸れると思います。私たちは、インパクトファクターに傾倒した不当な競争の議論から再現性の問題が浮上した過去を忘れず、歴史的経緯や文脈にそった評価としていただきたく存じます。

・本アンケートに関しては、特に後半半分はメタデータやビックデータなどの研究に関係している、あるいはその分野に関する知識が十分にあるものを対象としているように感じられる。（あるいは、募集段階で当該分野に関して理解していない人間以外を選定する必要が周知されていたのであれば、申し訳ないです。）

私はそういった観点での知識が殆どないことから、前半部分のデータの保存・公開に関しては理解し得るが、特に後半部分になるにつれ、回答の際に重要/重要でないではなく、そもそも質問の内容に関しての知識がなく、わからないというべき部分が多々あった。また、質問事項に関し、勘違いした回答を行っていることも強く懸念している。回答は用語を検索したうえで自分なりの解釈にて重要/重要でないを選択したが、できれば回答に”どちらでもない”に加え、”わからない”が欲しいと感じた。

・本機関においては各部より広くアンケートに協力することとなっているため、私のようにオープンデータなどに関して理解していないが、回答しているものもいるのではないかと考える。そのため、アンケートとして信頼性のおけるデータ取得を目的とするならば、理解度と併せて各項目の回答状況を分類するため、質問事項の中にメタデータ、オープンデータに関しての理解度や日頃の関わりに関する質問を設定し、分類した方がよいと思われる。

・上記の設問についてですが，あくまで自身のデータがどう利用されたか，どれくらい使われたかを知りたいだけで，それで何か評価されたい，と思っているわけではありません。自分が取得したデータを断りなく使われるのはとても嫌です。あくまでデータ取得したのが誰であるか，それさえ明らかにしてくれれば，あとはどうでもいいというのが本音です。仮に自分のデータを利用して，新しい薬ができて，どこかの企業が大儲けしたとしても，別に金銭をくれとは思いません。真の研究者は，現世での評価より，自身が死んだ後にも自分の見つけた事実（データ）が残ることを望んでいるものですし，それだけでいいのです。そもそも，「これまでの論文数・IF に加えて」とありますが，こういったことで研究者をランキングする習慣を止める方向に持っていくことはできないでしょうか。もしデータ公開での評価も更に行うということになると，質の悪いデータでも公開した方が「お得」みたいなことになりかねません。実際，これまで嘘でもいいから IF 高いところに掲載されたいと思う人がごまんといました。データの共有化，公開化を真に促進したいのであれば，まず研究者に点数をつけることを止めるべきです。そうすれば，誰かのため，将来の科学のために役立つデータを残したい，という人だけが，質の高いデータを公開するような時代になると思います。

・特定の研究分野や特許に絡むような研究は除くが，オープンサイエンスやオープンデータの意義や重要性を普及してほしい.

・理想と現実は異なり、自施設内の仲間内でさえ、データの提供の問題が現実的にはある。すなわち自分が何十時間もかけて準備したデータを、共同研究者だからという理由で提供し、提供先からは感謝の念を口頭で頂くが、実際にはデータ提供者であ

る自分には公となるコントリビューションは表明されず、上長からは、データ整理ばかりしていて業績がないという低い評価をされることが非常に多い。唯一この行いが世の中のためになっている正しい行いであるという自負のみがこの作業を支えている。しかしながら、同時に、競争として研究を行っている研究者の態度も非常に良く理解できるので、オープン化がそれらの問題をよく解決出来るのかどうか自信が無い。データを利用する場合は、データを利用しない場合よりも良い結果（たとえ誤用があったとしても、全体の知識としての蓄積はあるので）となると信じられるので、上記のような問題が少しでも解決されつつより広いデータ利用が可能になる事を望んでおります。

・1 次収集者・作成者の労苦・コスト・手間暇が適切に報われるような仕組みでない限り、うまくいかないだろう。

・データ共通に関しては、データを収集する立場の人への敬意と実利を確保し、バランスの良いシステムを作ることが必要不可欠です。人のデータをただ乗りして使えて当然、というような態度の研究者を放置した中で、共有のシステムを構築することは、全体の崩壊を招きます。現状は、データを収集することは自らの真実の探求のために必要であっても、それを人に使わせることは損はあっても得はありません。皆がハッピーになるような制度を願います。

・データを利用する研究者がデータ定義等を誤解して解析を行ってしまう危険性を回避するための教育が必要であるが、既存の研究者教育の中ではこの点が曖昧になされているのではないかと懸念される。また、オープンデータの流れを進めるためには、図書館員やデータキュレーターの存在が必須であるが、専門家としての司書の評価や待遇が低い現状を見るに、図書館員やデータキュレーターが活躍できる環境の構築に対して悲観的にならざるを得ない（自分自身は司書・図書館員・データキュレーターではないが、異なる専門分野の専門家として研究機関にそのような方が必要と感じている）。業績評価方法等も含め、体制整備が必要と考える。

・実際に非常に有用なデータで本来公開されるべきものが、おそらく企業の治験等の制限で公開されていない一方で、純粋にアカデミアのデータは無条件に提出が義務付けられている（少なくとも資金提供元からの要請、また論文のアクセプト要件として）。現実的に大規模データを扱える施設は限られており、恩恵を享受できる研究者は限られている。

大きなデータセットを持つ研究と、小さなデータセットを持つ研究を同列に扱うと、もともと大きなデータセットを持つ研究グループへの恩恵が大きいと思われる。

小さなデータセットを扱う研究者は、そのデータを十分に利用し終わるまで、公開を保留する権利を持っていても良いように思う。大きなデータセットを扱うためには、それなりの設備が必要で、もともと小さいデータセットを扱っている研究者が大きなデータセットを取り入れて解析するためには困難が大きい。一方、大きなデータセットを扱っている研究者が、小さなデータセットを吸収して、より大規模な解析を行うことは比較的容易であると思われる。

D. 考察

D.1 公開データの利用状況

本研究のアンケート回答率は 15.72%と低いが、これにはいくつかの理由が考えられる。参加研究機関に医療機関としての性格を持っているところがあること、また医療従事者としての意識が強い場合は、自身の取り扱う医療情報は機微な個人情報を多分に含み、最初からオープンデータ・オープンサイエンスの対象データとなりえず、ひいてそのようなデータのみ扱っている自身は回答対象となり得ないと認識した可能性が考えられる。今回のアンケート項目に入れていなかったが、診療にも常時従事しているかの属性を追加し、診療の従事の有無に

よる層別処理をすれば、上記のような仮説の検証ができると思われる。もう一つは、アンケート実施機関において新型コロナウイルス感染症対策の対応に追われている時期であり、特に医療従事者の立場を持つ者は回答のための時間を確保することが困難であったと思われる。

この回答率の低さによって生じるバイアスに関しては、未回答の多くが医療従事者であるという仮説に立つと、そもそもの原状の医療情報の多くはオープンデータとして公開しえないものであり、オープンデータの準備・公開にかかる検討への影響は少ないと思われる。それ以外に留意すべきと思われる箇所について述べる。

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(F1)について、上位 5 項目は先行研究[1]と同じであり、概ね同様な行動様式であることが確認された。また年代を通してもこの傾向は共通しており、公開したデータのパブリシティを効率的に獲得するために、主要なサーチエンジンに認識されるように公開したデータのメタデータを提供すること、論文や学術記事の参考文献に引用されるようなデータ引用に関する環境を整えることが望まれる。

公開データの入手先(F2)については、論文データベース(21.6%)、出版社(16.8&%)、政府統計(16.6%)、学術データアーカイブ(10.9%)と査読や厳密に手法が定められたプロセス等によってデータの品質が確認されているところからの入手が主流であった。先行研究[1]において 1 位であった「個人や研究室のウェブサイト」は本調査で 5 位になっており、本調査の対象となった研究者達のデータの信頼性を重視した姿勢が伺える。

経験年数が長い研究者はデータベースや研究者ネットワークを活用している傾向がある（表 4)。一方で、学術系 SNS やデータジャーナル等の最近出てきたツールの利用率は全年代を通して低調である。本調査に参加した国立研究機関の動向ではなく、その研究機関に所属している研究者の学術領域が、まだこれらのツールになじみのない分野であれば、これらのツールを通した訴求向上の取り組みの優先度は低いと思われる。データジャーナルについてはサーチエンジンに登録するメタデータやデータ引用の環境整備を進めていく過程で派生的に利用が増えていく可能性がある。すなわち、現状の査読やオーソライズされたシステムの仕組みの延長上に公開したデータを効率的に流通する枠組みを整えることが有用な取り組みであろう。

D. 2データ保持状況

研究データを保持するモチベーション(G1)については先行研究[2]では上位 4 位に共通した傾向を持っていた。さらに、他はないデータであること、学際的な連携を推進することについては先行研究[2]よりも重要視されていた。一方で経済的価値は相対的に重要視しない傾向は共通している。将来における再検証可能性確保の理念的な有用性は認めつつも、実験条件等の違いにより必ずしも有用ではないことが指摘されている。この課題を克服するにはデータが時代を超えて再利用可能性を備えるように整備する必要があろう。モチベーション以前にデータを保管することは義務だという指摘が複数ある一方で、長期間に渡る管理の手間・懸念の指摘も見られた。本調査の結果を踏まえると、将来における検証、再検証ができるように可及的にデータの構造化・標準規格の準拠を推進し、研究者から保管の負担と責任を減じるような、長期間に渡ってセキュアに保管するプラットフォームの開発と提供が有効な施策であると考えられる。

研究データの保持に関する課題(G3)については、全ての項目に対して懸念が寄せられたが、「全体的な賛成率」と「非常に重要であるとした割合」のそれぞれ上位 5 位で共通していたものは、「データ管理を任せていた人が期待に添えなくなる」「持続可能な

ソフトウェア・ハードウェアのサポートの欠如」「プロジェクトあるいは組織によるデータの管理が終了する可能性」であった。個人従属性の高い業務は組織・予算の事情に対して脆弱である。我が国では情報管理に関する専門家は極めて少なく、またその専門家を長期間に渡って安定的に雇用する体制に乏しい。長期間に渡って信頼性のある形でデータを保存・管理する取り組みを個々の機関で行うことは困難が伴う。データ管理について責任を負うシステムあるいは組織の外部化・委託によって継続性を担保する取り組みが必要であろう。

また自由記述欄の回答においても、データの品質管理の概念の不足が指摘されており、第三者がデータを利用するためのメタデータの管理不足や、フォーマットの陳旧化によってデータシェアを目的としながら、利用機会が少なくデータアーカイブに留まる。公的資金による支援の形式や持続可能性のあるビジネスモデルを国が持つことの重要性が指摘されている。データ公開や保存に関して倫理指針、インフォームドコンセントに関するルールや匿名加工の具体的基準に欠けていること。そしてそれが倫理委員会での個別検討を強いられ、データ公開にあたって更なる倫理委員会の審査対応の労力が増えることの懸念がある。データの保存義務・期間について指針・規定・助成機関の方針の衝突があり、消去法的に長期間にあわせて保管する中でデータの破棄がなされずに管理責任も不明になるというリスクがあることが指摘されていた。

現状の環境ではデータ管理とメタデータの作成を研究者に担わせることは負担を増やすことにつながる。データ管理とメタデータの作成を専門に担当する人材を雇用し、研究者を支援させることが望ましいと考える。一方で、データを一番理解し、またデータそのものを生成するのは研究者であるので、今後のデータ作成は可能な限り標準規格に準拠したもの、あるいは広く使われておりデータに関する情報も含まれているようなデータ形式を採用するように働きかけるのが望ましい。研究者から独自形式のデータを受け取り、データ管理者の手で標準形式に変換するのは、さらなる労力を伴うだけではなく、本来のデータから変容したものになるリスクがありうる。データ公開の判断基準、匿名加工の具体的基準、データ公開を前提とした倫理指針やインフォームドコンセントのガイドラインの整備も必要である。データの保管期限については個別判断となるが、低廉に運用可能であり、長期的に安全性が保証されるシステムがあれば、長期間に渡るデータの保存に関するリスクを軽減させられるだろう。

データ保持に関する知識を増やす有用な方法(G5)では、有用の評価が上回ったのは「デジタルデータ保持に関するガイドライン・マニュアルの開発(36%)」のみであった。先行研究[2]では 4 項目ともに有用性が上回っていたことと対照的である。自由記述では、個人のスキル向上も大事であるが、研究機関・国レベルでの大きな枠組みで検討することの重要性、研究助成機関がデータ保持の為の研究分野を開拓すること、データ公開に係る手続き・業務の煩雑さ、データ利活用の意義や重要性を啓発し、研究者にインセンティブがあれば発展的に解決していくという指摘があった。研究データの保管は直面している課題であり直近のニーズとして捉えられるが、具体的なトレーニングやプラットフォームについては第三者の専門家マターとしてお願いしたいという意向があると思われる。国・研究機関としてデータ保存・公開にかかわるプラットフォームを開発し、研究者として最低限なすべきこととして、自身の研究データの保全ができるようなガイドラインを提供する。その後のデータ公開については専門家の支援を受けられるようにすれば、オープンサイエンスの実践が無理なく取り組まれるようになると思われる。また研究助成機関にはオープンサイエンスに特化した分野項目を設けて研究を推進するよう提案すること

が考えられる。

D.3データの保持対象

「そのままの利用は難しいがデータに関する説明を付けて公開すれば利用される可能性がある。」、「データフォーマット等互換性の問題が解決されたら可能性がある。」が確認され、自由記述においてもデータの誤用への懸念、そして誤用を防ぐための必要十分な情報提供がされることが重要であるが、その情報提供の準備に割く人的資源の不足が指摘されている。異分野での利用はオープンマインドではあるが、現実的問題としてそれが成立するための環境整備に回す余力がない現場の実情が伺える。

他者がデータを利用する際の公正な条件(J10)について、過半数を超えた上位回答は、研究者による正式な引用が必要、データの相互共有に関する相互合意、データ提供者への謝辞、データ提供者による承認、プロジェクトで協力する機会の提供であった。小野[7]、Tenopir[5]と比較して、先行研究では相対的に優先度が低かった「データを活用する研究活動についてデータ提供者あるいは研究資金提供機関の正式な承認が必要である」とした回答が上位に来た点において特徴的であった。第 5 期科学技術基本計画[12]では「オープンサイエンスとは、オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である。」と定義されている。すなわち、オープンサイエンスは研究論文を中心としたオープン化を目指すオープンアクセスと研究データのオープン化(オープンリサーチデータ)を含む概念とし、研究成果をよりオープンにして利活用を推進させることにより、研究を加速することが期待されている[13]。そして、Budapest Open Access Initiative(BOAI)によれば、誰もが自由にアクセスでき,かつ自由に再利用できることがオープンアクセスの要件[14]とされている。但し、単にネット上に公開するだけでなく、二次利用には原則として権利者の許諾が必要であるという著作権の問題を解決するために利用条件の意志表示を示す必要があり[15]、それを簡易に実現するためのツールとしてクリエイティブコモンズ(Creative Commons :CC)[16]が多用されている。CC ライセンスで公開されたものは、CC ライセンスに準拠した利用をする限り、データ公開元に「利用の承認」を都度取り付ける必要はなく、それが迅速なデータ利活用につながっている。このような時流の中、「データ提供者あるいは研究資金提供機関の正式な承認が必要である」というスタンスは、どこから来るのかを丁寧に検討してフォローする必要があると思われる。

データを公開する際のライセンス形態(J11,J12)について、本調査は過半数を占めた上位順に CC-表示、CC-改変禁止、犯罪・違反時に利用禁止、CC-非営利、犯罪・違反時に罰則であった。小野[7]の研究と比較して、CC-改変禁止の優先度が高いこと、CC-継承が低いことが特徴的であった。CC-改変禁止の支持、CC-継承の不支持については、元データから様態を変更される余地を許容することにより、他項目で確認されている、データの誤用・悪用を懸念していること、営利目的利用に対して否定的な意見がある背景から生じていると考えられる。オープンデータの原理的な取り組みについて理解を示しつつも、データの不適切な利用による科学の発展を結果的に阻害することの懸念が衝突していると思われる。

データを公開した時の関心のある指標(J13)について、過半数を超えた意見は、引用数(85%)、閲覧数(71%)、ダウンロード数(65%)であった。データ作成に対する敬意、業績の評価のために引用を希望する意見が多いと思われる。Impact Factor に続くねじれた評価につながることを懸念する声も一考に値する。ただ、これまでの評価項目よりデータ作成者に対する評価、支援が圧倒的に足りていないことが指摘されており、長期展望的には科学技術の発展の足枷となることが懸念されるのであれば、データの引用にかかる環境を整えてデータ作成者に対

する評価指標を確立することは重要な施策ではないかと思われる。その上で、ねじれた評価に濫用されないように我々科学者内部の努力も多いに求められていると考える。

第三者に提供できない理由(J14)について、小野ら[7]の全分野の傾向と同様、データ公開にあたって、時間・予算上の制約、メタデータ作成等の技術がない等の指摘があった。一方、小野ら[7]での医学・健康領域の約 10 名による回答とは様相が相当異なっていた。本調査での重要な懸念事項として浮上しているデータの誤用や、法律・規則上の課題への回答が見られないなど、対照的な結果となっている。本調査の方が医学・健康領域の人数が多く網羅的に回収することを試みていることから、先行研究では出てこなかった懸念を浮き彫りにしたのではないかと考える。Tenopir [5]の設問は「電子的にデータを提供できない理由」であるので、趣旨は異なるが、やはり時間・予算の欠如、データを公開する権利がない、データを公開する場所がない、標準規格の欠如、が上位の課題として提示されていたので、最優先で検討すべき課題は「時間・予算の確保」に対して支援することと設定してよいと思われる。

D. 4データに関する技能

今後受けたいトレーニング等(K1)として、上位3位はデータの安全な管理、適切なデータ形式、知的財産やライセンスであり、池内らの報告[1]の上位3位と順番は違えども同内容であった。また、第三者による支援が必要な物(K2)として、上位回答に「適切なデータ形式の変換」、「知的財産権やライセンス」、「適切なリポジトリの選択」、「リポジトリへのデータ手続き」であった。別項で、研究者の管理負担を減らすためにデータ管理者の専門家の雇用、データ保管・公開の為のクラウドプラットフォームの採用を提案している。それらの提案が採用されるという前提に立てば、専門家に適切なリポジトリのキュレーション、適切なデータ形式の変換、手続きの支援を期待できるものとして、それでも研究者がおさえておくべき内容にトレーニングテーマを絞ることが考えられる。すなわち、データの作成者である研究者がおさえるべき適切なデータ形式と知的財産・ライセンスと研究活動にも直接関わるテーマでのトレーニング開発から着手することが望ましいと考える。

D. 5データ公開に関する機関の個別傾向

データポリシーにかかる提言を検討するにあたり、組織横断的なポリシーを策定すべきかを検討するために、データポリシーの策定にあたっては、データの公開にあたってのデータの整備方針や、データ提供に関する条件、ライセンスが組織方針に影響を受けると思われるため、「セクション J: §研究データの提供状況」における3つの設問に関して各組織の回答状況をドリルダウン分析した。設問 J8「データを公開した場合、異分野の研究によって利用できると思われますか？」について、組織ごとに分析したものを表 16 に示す。「データに関する説明をつけて理解できるような形で公開すれば他分野でも利用される可能性がある」「データフォーマットなど互換性の問題が解決されれば利用出来る可能性がある」において回答状況が特徴的に分かれていることが確認された。データ公開における説明の付与の必要性はデータ公開にあたっての準備工程を増加させるし、データフォーマットの互換性については一研究者で対応できる範囲を越えている可能性があるため、この分野において対応が必要と認識している組織においては公開データの準備に関わるポリシーについて配慮が必要と思われる。

表 16 組織毎のデータ利用可能性の志向

研究機関	説明付与	互換性	ニーズ	不明
国立保健医療科学院	43.8%	43.8%	15.6%	25.0%
国立医薬品食品衛生研究所	48.0%	36.0%	16.0%	32.0%
国立感染症研究所	40.6%	21.9%	9.4%	25.0%
国立研究開発法人医薬基盤・健康・栄養研究所	46.9%	34.4%	0.0%	25.0%
国立研究開発法人国立がん研究センター	43.3%	40.7%	14.0%	26.0%
国立研究開発法人国立国際医療研究センター	23.5%	23.5%	11.8%	11.8%
国立研究開発法人国立循環器病研究センター	50.0%	50.0%	12.5%	25.0%
国立研究開発法人国立精神・神経医療研究センター	57.4%	54.1%	14.8%	27.9%
国立研究開発法人国立長寿医療研究センター	50.0%	31.3%	6.3%	6.3%
国立社会保障・人口問題研究所	45.5%	45.5%	9.1%	45.5%
国立障害リハビリテーションセンター	50.0%	44.4%	19.4%	33.3%
独立行政法人労働者健康安全機構	50.0%	50.0%	22.2%	33.3%
独立行政法人国立病院機構	25.0%	18.8%	6.3%	18.8%
独立行政法人国立重度知的障害者総合施設のぞみの園	66.7%	33.3%	16.7%	50.0%

設問「J10 他人が貴方(データ提供者)のデータを使用する場合を使用するに 際して、公正な

表 17 各組織のデータ利用条件に関する志向

研究機関	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
国立保健医療科学院	28.1%	40.6%	59.4%	50.0%	40.6%	28.1%	15.6%	25.0%	31.3%	34.4%	21.9%	46.9%	40.6%
国立医薬品食品衛生研究所	42.7%	56.0%	84.0%	65.3%	66.7%	48.0%	24.0%	38.7%	38.7%	36.0%	16.0%	64.0%	41.3%
国立感染症研究所	21.9%	40.6%	65.6%	46.9%	37.5%	34.4%	28.1%	18.8%	12.5%	21.9%	12.5%	50.0%	21.9%
国立研究開発法人医薬基盤・健康・栄養研究所	34.4%	46.9%	65.6%	62.5%	50.0%	40.6%	31.3%	28.1%	21.9%	28.1%	21.9%	59.4%	43.8%
国立研究開発法人国立がん研究センター	42.7%	52.0%	65.3%	57.3%	47.3%	38.7%	32.7%	28.7%	29.3%	34.7%	31.3%	60.0%	38.7%
国立研究開発法人国立国際医療研究センター	41.2%	50.0%	58.8%	35.3%	35.3%	35.3%	20.6%	32.4%	23.5%	38.2%	35.3%	50.0%	35.3%
国立研究開発法人国立循環器病研究センター	62.5%	62.5%	87.5%	75.0%	50.0%	37.5%	0.0%	37.5%	37.5%	50.0%	37.5%	75.0%	37.5%
国立研究開発法人国立精神・神経医療研究センター	59.0%	68.9%	82.0%	70.5%	68.9%	55.7%	45.9%	39.3%	31.1%	49.2%	37.7%	78.7%	60.7%
国立研究開発法人国立長寿医療研究センター	37.5%	31.3%	50.0%	50.0%	43.8%	25.0%	18.8%	18.8%	18.8%	31.3%	18.8%	50.0%	25.0%
国立社会保障・人口問題研究所	27.3%	90.9%	72.7%	54.5%	18.2%	45.5%	27.3%	27.3%	36.4%	54.5%	27.3%	81.8%	72.7%
国立障害者リハビリテーションセンター	25.0%	44.4%	86.1%	61.1%	55.6%	41.7%	30.6%	30.6%	44.4%	50.0%	27.8%	66.7%	58.3%
独立行政法人労働者健康安全機構	33.3%	66.7%	83.3%	66.7%	38.9%	50.0%	33.3%	38.9%	55.6%	61.1%	38.9%	61.1%	66.7%
独立行政法人国立病院機構	31.3%	56.3%	62.5%	25.0%	43.8%	37.5%	18.8%	37.5%	25.0%	31.3%	37.5%	56.3%	50.0%
独立行政法人国立重度知的障害者総合施設のぞみの園	0.0%	50.0%	83.3%	66.7%	66.7%	16.7%	50.0%	66.7%	66.7%	66.7%	33.3%	100.0%	83.3%

条件はどのようなものであるべきかのお考えについてお尋ねいたします。」に関して各組織の回答状況を表 17 に示す。個々の項目の意味は別添のアンケート文面を参照されたい。組織ごとの回答の分散が大きかった順に、Q13、Q1、Q12 であった。設問 Q1「当該データを使用した論文の共著者としてもらう」についてはオープンデータとして引用して貰えるように、引用方法が明確になるようなメタデータの整備をすることに力をいれることになるだろう。一方、設問 Q13「データ提供者は利用するデータの使用目的等に関する同意書を受け取る必要がある」設問 Q12「データの相互共有に関する相互合意が必要である」とする方針は、可及的に流通性を高めることを志向しているオープンデータや Creative Commons の基本的方針とは方向性が異なるものになる可能性があり、この部分は組織の性質にも関わると思われる。

設問 J11「データを公開する際のライセンス形態で相応しいと思われるものについて選択して下さい」に関して各組織の回答状況を表 18 に示す。各組織の回答率の分散が大きかったものから、犯罪・違反時に利用禁止、CC-改変禁止、CC-非営利となった。Creative Commons における改変禁止とは、元データの改変を禁止することではなく、改変して新しいものを作ったときに改変された資料を頒布する行為を禁止することである。いくつかの側面が考えられる。データそのものを改変することを改竄行為と見做すが故の回答、あるいは派生データが多く流通すると一次データが不明になる可能性があることから、可能な限り一次データからの引用・利用を求めるという考え方からの回答も考えられる。ケース・バイ・ケースではあるが、派生データの配布を抑制することについて組織としてのポリシーが分かれる可能性がある。また CC-非営利は、原作者のクレジットを表示し、かつ非営利目的であることを条件に改変したり再配布したりできる（改変ができるかは、改変禁止の属性を追加するかで決まる）とするものである。非営利といっても営利用出来ないわけではなく、個別に原作者から営利用を許可するライセンスを交付すればよい。このライセンスの考え方についても組織によって方針が異なることが伺える。

表 18 各組織の公開ライセンスに関する志向

研究機関	CC-表示	CC-継承	CC-非営利	CC-改変禁止	利用禁止	謝辞
国立保健医療科学院	43.8%	21.9%	37.5%	46.9%	43.8%	25.0%
国立医薬品食品衛生研究所	66.7%	36.0%	58.7%	58.7%	49.3%	42.7%
国立感染症研究所	56.3%	25.0%	43.8%	53.1%	40.6%	18.8%
国立研究開発法人医薬基盤・健康・栄養研究所	50.0%	43.8%	46.9%	43.8%	40.6%	40.6%
国立研究開発法人国立がん研究センター	52.7%	23.3%	44.0%	46.7%	48.0%	39.3%
国立研究開発法人国立国際医療研究センター	44.1%	23.5%	41.2%	38.2%	44.1%	41.2%
国立研究開発法人国立循環器病研究センター	62.5%	50.0%	75.0%	75.0%	75.0%	62.5%
国立研究開発法人国立精神・神経医療研究センター	70.5%	41.0%	57.4%	70.5%	70.5%	55.7%
国立研究開発法人国立長寿医療研究センター	37.5%	18.8%	31.3%	43.8%	43.8%	31.3%
国立社会保障・人口問題研究所	54.5%	18.2%	63.6%	54.5%	81.8%	36.4%
国立障害者リハビリテーションセンター	69.4%	30.6%	63.9%	77.8%	61.1%	41.7%
独立行政法人労働者健康安全機構	55.6%	38.9%	66.7%	66.7%	38.9%	33.3%
独立行政法人国立病院機構	37.5%	25.0%	37.5%	31.3%	25.0%	25.0%
独立行政法人国立重度知的障害者総合施設のぞみの園	83.3%	33.3%	83.3%	83.3%	100.0%	50.0%

以上のことから、オープンデータの公開にあたっての方針について組織の事情が異なること、また組織内において複数分野の研究者が在籍しており、研究分野ごとに個別事情があることも踏まえて、組織横断的なポリシーを策定することは推奨されないと考える。

E. 結論（提言）

本調査の結果を踏まえると、一律のデータポリシーを策定するのではなく、データポリシーは各研究機関が、その独自性を踏まえて作成することが望ましい。その上で、全体的にオープンサイエンスにむけて取り組むべき事項を以降に提案したい。

E. 1 必要な人材像

E. 1. 1 公開データのパブリシティを向上させるために、既存の検索エンジン、論文データベースから認識されるようなメタデータの作成ができる環境を整備すること。またメタデータの作成を支援する人材を配置すること。

E. 1. 2 データ公開に伴う安全性や匿名加

工処理等の検討や支援ができる専門家の養成と配置を検討すること。

E.1.3 メタデータの作成を支援できる人材を増やすために、メタデータ作成に関する教育を提供するコンテンツを開発すること。

E.2研究組織

E.2.1 長期間に渡り安全に公開データを保存・管理するプラットフォームを構築し、研究者・研究機関に提供すること。事業継続性を担保するためにプラットフォームの運用コストは低廉なものとし、また規約・制度を整備することでプラットフォームにデータを預託した研究者の管理責任を減じること。研究機関によって、プラットフォームの運用に関わる情報システム担当者を設置できるかの状況は異なるため、プラットフォームの運用は研究機関横断的あるいは研究機関ごとに柔軟に対応できるようなものであること。すなわち、プラットフォームは研究機関内部（オンプレミス）に設置するのではなく、クラウドをベースとして外部に管理を委託できるものが望ましい。但し、各研究機関の研究分野の特殊性も考慮し、柔軟に設定を対応できるようなものであること。

E.2.2 オープンアクセス、オープンデータを標榜しているジャーナルへの投稿を推薦し、研究活動のパブリシティ向上に努める。本来の研究活動である論文投稿、そして論文の補遺としてデータを公開する経験を通して、リポジトリでのデータ公開までの道程を無理なく進めるように支援していく。

E.2.3 データ作成・公開の業績を評価するメトリクスを開発すること。これまで光が当てられることの少なかったデータ作成・管理に関わる者の適正な業績評価につながることを志向する。しかし、論文数・Impact Factor を用いた評価では、研究者が多い学術領域、大規模な研究機関で共同研究者に恵まれた者が業績を獲得しやすく研究資源の多寡が研究者の評価に影響をあたえかねないという様々なねじれが生じてきた過去の反省を踏まえ、取扱いには慎重を期したい。

E.2.4 オープン＆クローズ戦略が確立されていない研究機関は、オープンデータポリシーの検討に伴い、既存の規則の活用に加え、職務発明規程や就業規則の修正・追補を検討すること。

E.2.5 地方自治体等提供元の意向によりオープンデータとして開示出来ない旨の事例が散見されたことについては、提供元の事情を配慮しつつ可能な範囲で提供・オープン化を依頼すること。例えば、現行の個人情報保護法においては匿名化されたデータは個人情報にあたらないこと、データ公開にあたって匿名加工を施すなど安全性を確保すること、オープンサイエンスの意義を説明することにより、ご理解を頂くよう啓発活動を企画すること。

E.2.6 倫理委員会及び審査員において、データの公開を前提とした研究計画であるかを確認し、必要な工程、施策について助言すること。

E.3国・省庁の対応

E.3.1 オープンデータによる情報価値の共有・研究推進がもたらす社会的な意義を国民に啓発する活動を継続すること

E.3.2 国・研究助成機関に、オープンサイエンスを推進するべく、新規の研究領域・分野を設けて公募するよう提案する。具体的には、下記のテーマについて考慮することを求める。(1)研究データの安全な保管・公開を実現するプラットフォーム、(2)研究データとメタデータの標準化にかかる検討、(3)標準規格に準拠する為のツール開発、(4)データ公開にむけた匿名加工処理の具体的基準・手順開発

E.3.3 メタデータの作成を支援する専門家、データ公開に伴う安全性や匿名加工処理等の検討や支援ができる専門家を研究機関が継続的に雇用できるように予算を確保すること。

E.3.4 不適切なデータ解釈による研究論

文の粗造を懸念する声については、査読を通したコントロールという研究者集団による自助努力を要請するものとする。しかし、その大前提として、適切なデータ利用を促すメタデータと標準規格の準拠を実現するための人的資源、予算の確保が大前提である。また査読についても大きなリソースを割くために、オープンデータを扱いうる学術全体の底上げが必要である。すなわち、適切な追加の助成措置なくして研究者の自助努力を要請するのみならば、これまで以上の研究環境の崩壊を招きかねないことに留意されたい。

F. 謝辞

本研究は、厚生労働科学特別研究事業「厚生労働分野のオープンサイエンス推進に向けたデータポリシー策定に資する研究(201906008A)」の助成にもとづき行われた。

本研究は、国立保険医療科学院の研究倫理専門委員会にて、課題名「厚生労働分野のオープンサイエンス推進に向けたデータポリシー策定に資する研究」(NIPH-IBRA#12265)として承認された。またアンケートに参加した各機関においても、倫理審査を実施し、承認後にアンケートを実施した。

新型コロナウイルス感染症対策の対応を行っている中、貴重な時間を割いて委細な意見を自由文章形式にてもご回答頂いた研究者の皆様方にこの場を借りて御礼を申し上げます。

G. 参考文献

1. 池内有為, 林和弘, and 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*. 2017, 科学技術・学術政策研究所.
2. Kuipers, T. and J. van der Hoeven, *Insight into digital preservation of research output in Europe.* Survey Report, 2009.
3. Likert, R., *A technique for the measurement of attitudes.* Archives of psychology, 1932.
4. 国立研究開発法人科学技術振興機構. *e-Rad における研究分野一覧*. 2019; Available from: https://www.jst.go.jp/coi/download/file/keyword.pdf.
5. Tenopir, C., et al., *Data sharing by scientists: practices and perceptions.* PloS one, 2011. **6**(6): p. e21101.
6. Ferguson, L., *How and why researchers share data (and why they don't).* Discover the future research. The Wiley Network, John Wiley and Sons, Hoboken, New Jersey, USA [online]: Available from hub.wiley. com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont, 2014.
7. 小野雅史, 小池俊雄, and 柴崎亮介, *地球環境情報分野における研究データ共有に関する意識調査: 研究現場の実態.* 情報管理, 2016. 59(8): p. 514-525.
8. Schmitz, C., *LimeSurvey: An open source survey tool.* LimeSurvey Project Hamburg, Germany. URL http://www. limesurvey. org, 2012.
9. Team, R.C., *R: A language and environment for statistical computing.* 2013.
10. Bryer, J., K. Speerschneider, and M.J. Bryer, *Package 'likert'.* Analysis and Visualization Likert Items. CRAN, 2016.
11. Berners-Lee, T., *Five star open data.* 2009.
12. 内閣府. *第5期科学技術基本計画.* 2016; Available from:

https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf.
13. 林和弘, *オープンアクセスとオープンサイエンスの最近の動向: ビジョンと喫緊の課題.* 表面科学, 2016. 37(6): p. 258-262.
14. Brown, P.O., et al., *Bethesda statement on open access publishing.* 2003.
15. 水野祐, *オープンアクセスとクリエイティブ・コモンズ採用における注意点: 開かれた研究成果の利活用のために.* 情報管理, 2016. 59(7): p. 433-440.
16. Lessig, L., *The creative commons.* Mont. L. Rev., 2004. 65: p. 1.