

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

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関する研究

令和元年度 総括・分担研究報告書

研究代表者 木村 映善

令和2(2020)年 7月

目 次

I. 総括研究報告	
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	----- 1
木村映善	
(資料) オープンサイエンスに関する調査のアンケート調査用紙	
II. 分担研究報告	
1. 機微なデータの公開と共有	----- 58
ANDS ガイドからの機微なデータの公表と共有に関する主要メッセージ	
木村映善	
(資料1) 機微なデータの発表と共有にむけての判断フローチャート	
(資料2) 機微なデータの発表と共有にむけての判断フローチャート (国内制度対応版)	
2. 国立研究開発法人におけるデータポリシー策定にあたり検討すべき事項の整理 ～「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」を羅 針盤として～	----- 74
木村映善	
III. 研究成果の刊行に関する一覧表	----- 90

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)

高橋 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（紙資料であればスキャンして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)については、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

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、

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

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。

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		Total
	男性	女性	
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
検索エンジン (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 (Elsevier, 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 (Data Citation Index など)	125	7.5%
政府・国際機関・出版社などの 広報・ニュースレター	96	5.7%
学術系 SNS (Mendeley, Research Gate 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 (Facebook, Twitter 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					合計
	5年未満	5~10年	11~15年	16~20年	21年以上	
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 (Data Citation Index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 (Elsevier, 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・ニュースレター	11	10	20	19	36	96
ブログや一般的なSNS (Facebook, Twitter など)	7	7	9	5	10	38
学術系SNS (Mendeley, Research Gate 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。経験年数を通じた検索手段の分布もほぼ同等であったが、経験年数が高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。

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、ここで尋ね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 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 (figshare, zenodo など)	2	0.2%
回答者数	1276	100.0%

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					合計
	5年未満	5~10年	11~15年	16~20年	21年以上	
論文データベース (Pubmed Central, J-Stage, 医中誌)	19	38	56	46	117	276
出版社 (Elsevier, Wiley-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 (Mendeley, Research Gate 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 (figshare, 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(新しい研究は既存の知識にもとづいて実施可能)(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、研究結果は公的財産のものであるため(84%)、他に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情動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

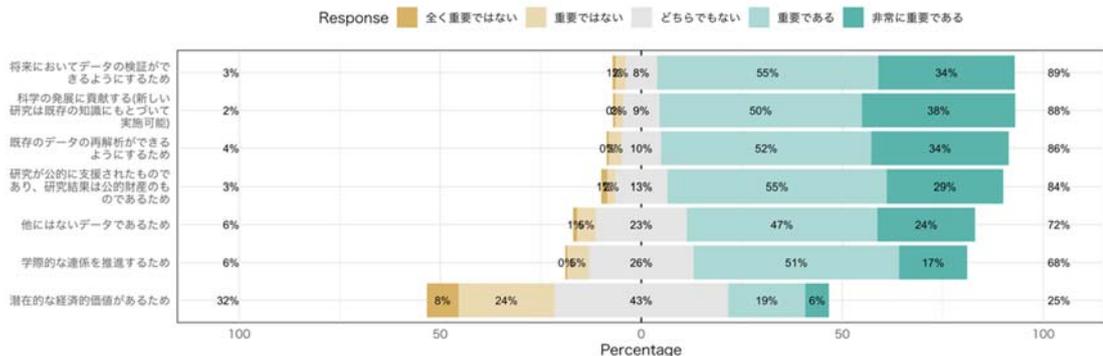


図 2 データを保持するモチベーション(Q G1)

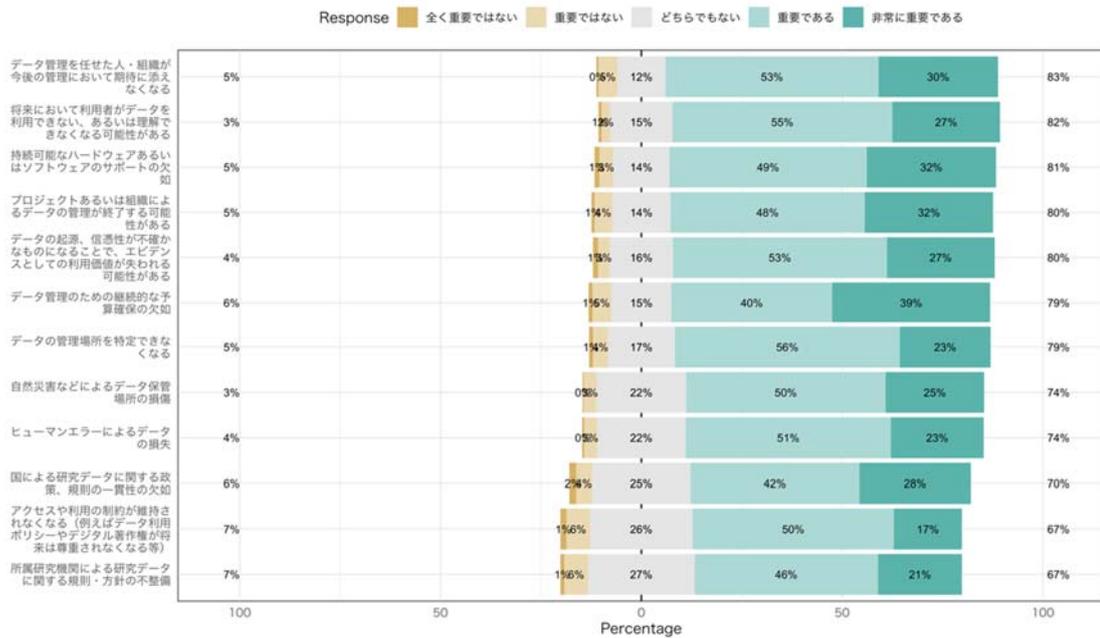


図 3 データの保持に関する課題 (Q G3)

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だと思います。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のでしょうか？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られ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]

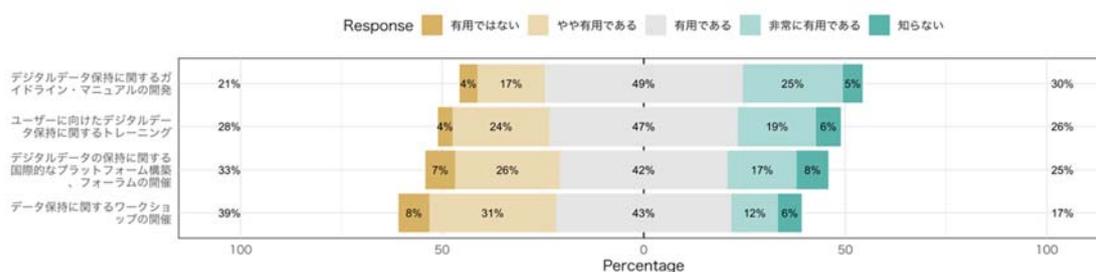
・ データの価値を素早く正確に評価することは、多面的な側面から難しい。

・ 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られる。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くますが、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ーム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) とし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の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、チャット

表 8 研究者が所属する研究分野 (Q H2)

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

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

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%

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。

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%

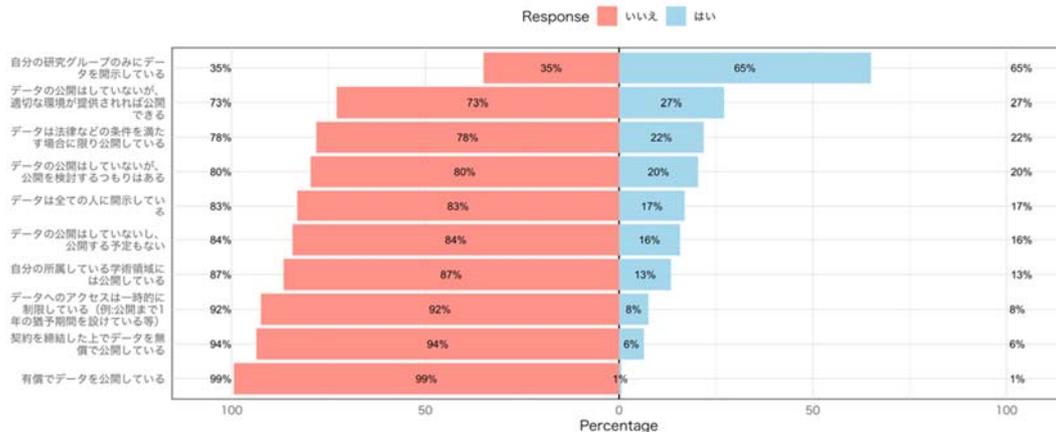


図 5 データの公開状況 (Q II)

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 II）」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

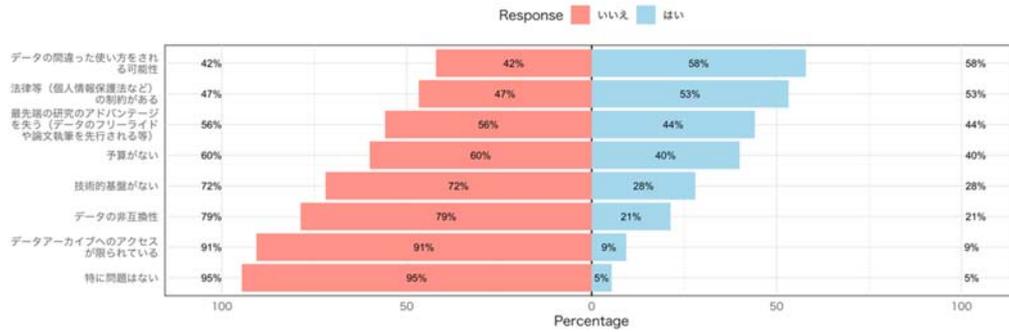


図 6 データ公開の障壁 (Q I2)

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。

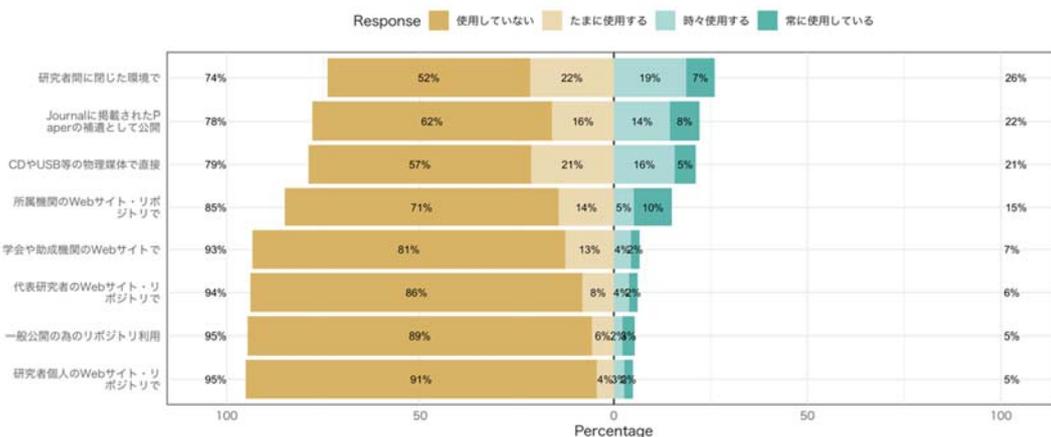
自由記述欄の回答を以下に示す。

- ・ 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通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場所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

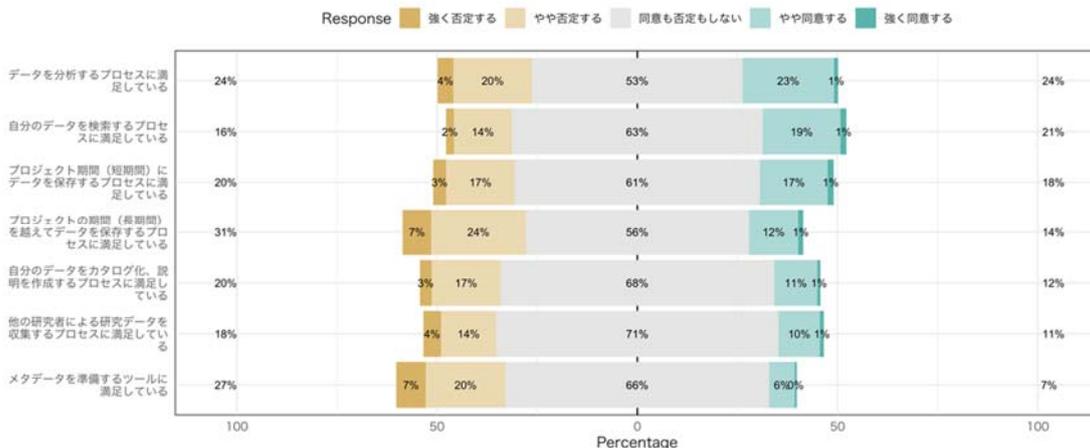
C. 6.2 データ収集・準備プロセスへの満足度

「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)



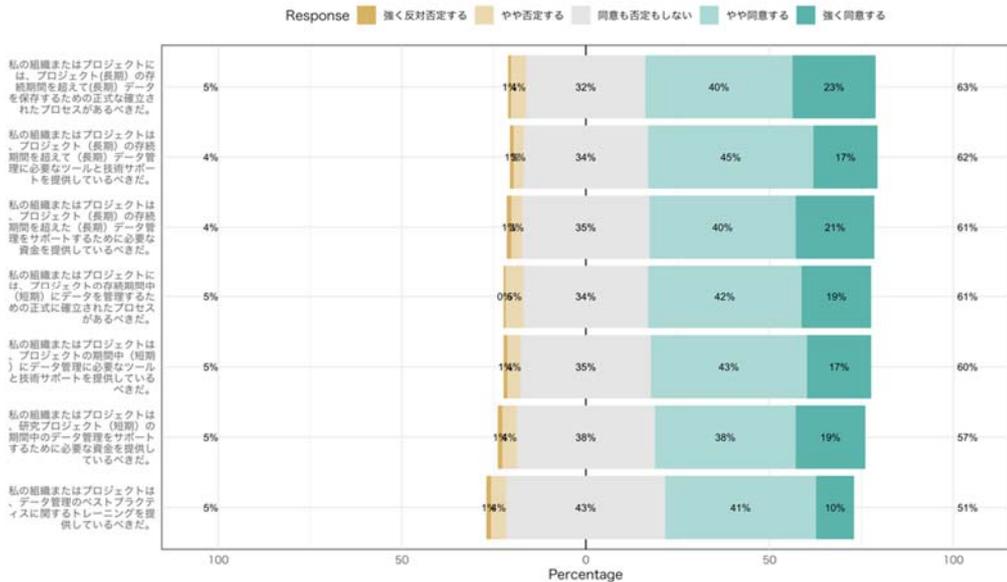


図 9 組織・プロジェクトのデータへの関与(Q14)

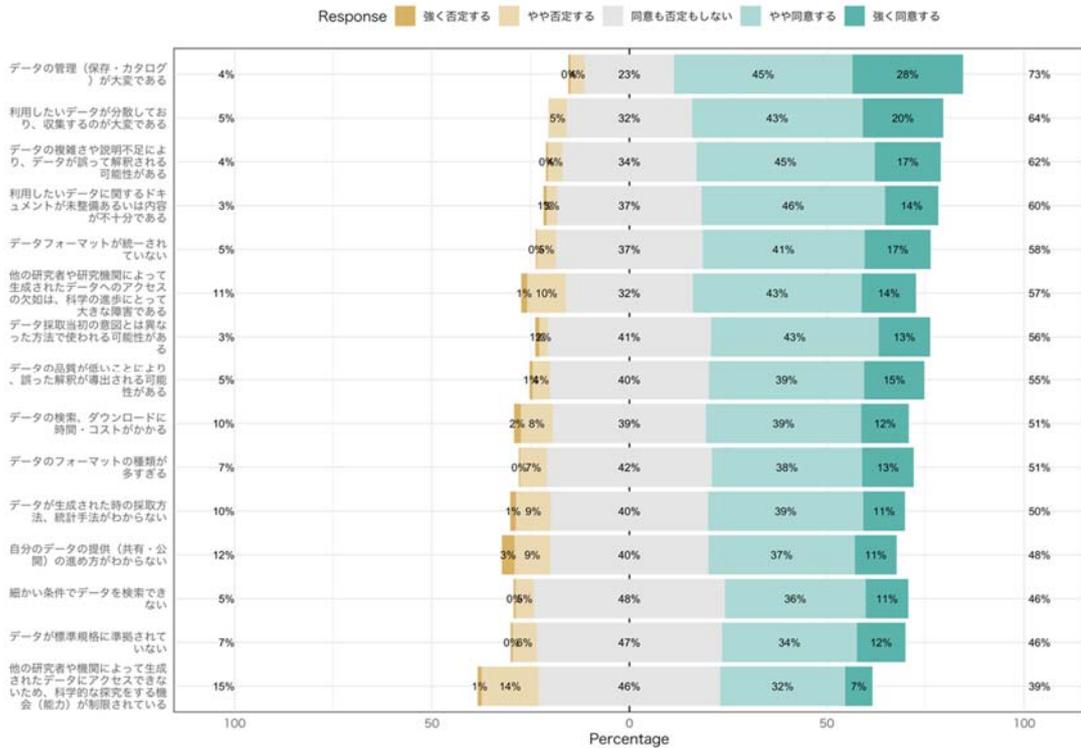


図 10 データ利用に関する課題・懸念(Q15)

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する課題、懸念事項について意見をお伺いします。(Q15)」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する課題、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それ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りにはなっていない。SS-MIX 2の取り組みも十分生かされていないように見える。

- ・ 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いますが、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、正確性の高いデータが、正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

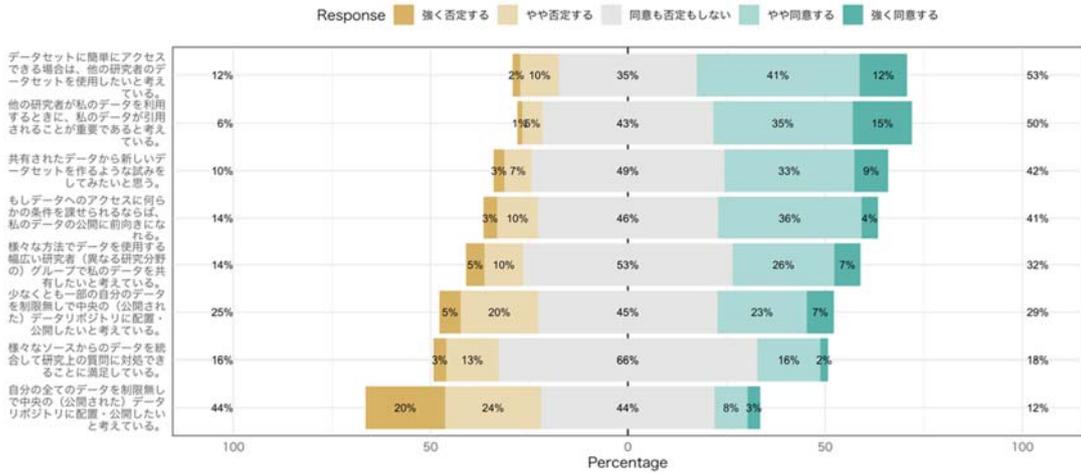


図 11 データ共有に関する意向 (Q J7)

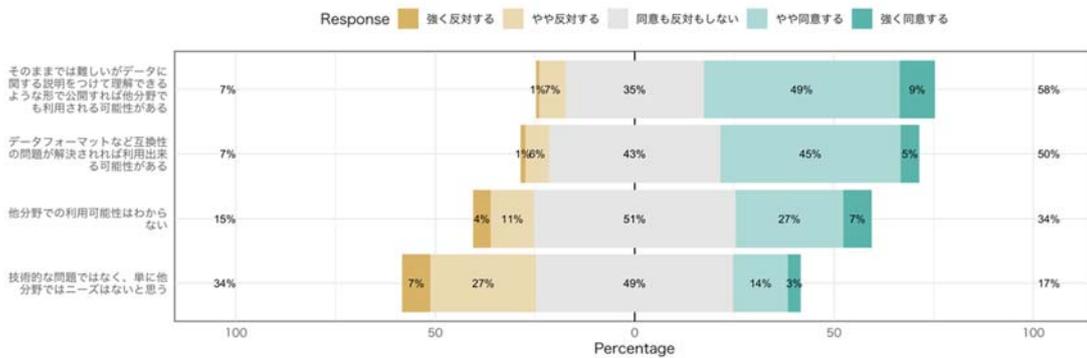


図 12 異分野の研究における利用可能性 (Q J8)

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、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、使われ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

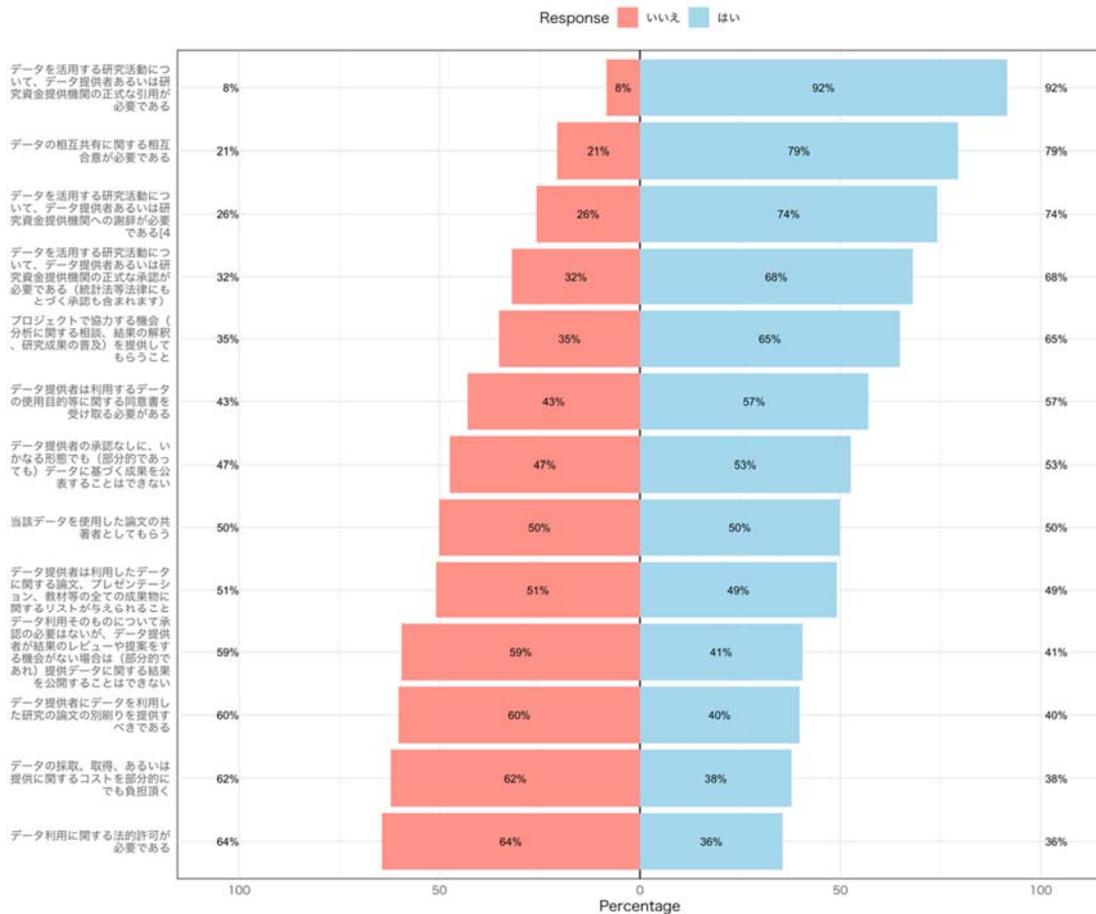


図 13 データ使用時の公正な条件(Q J10)

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

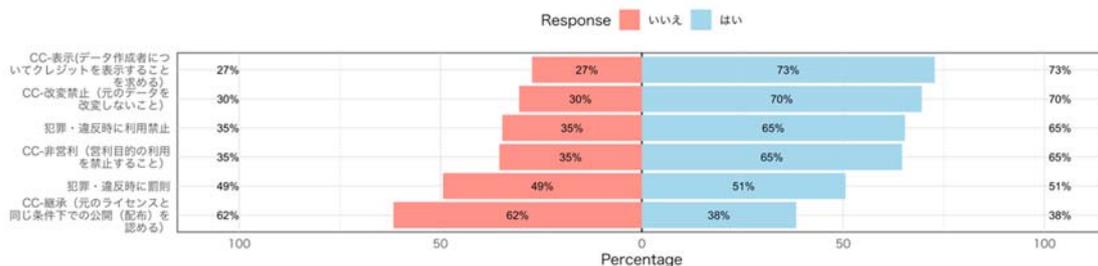


図 14 データ公開のライセンス (Q J11)

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

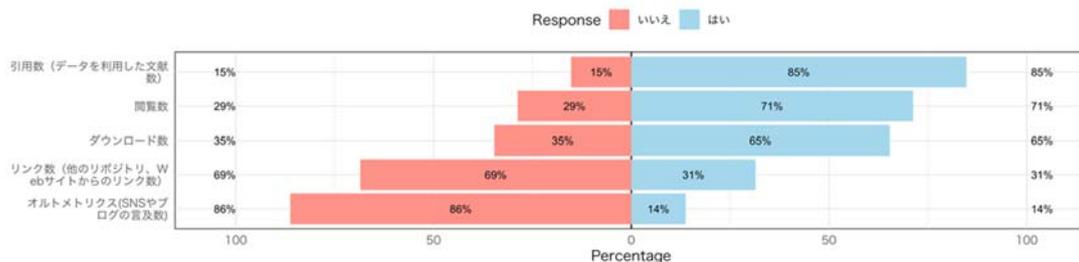


図 15 データ利用に関するメトリクス (Q J13)

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ついての回答です
- ・ 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る

・ 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

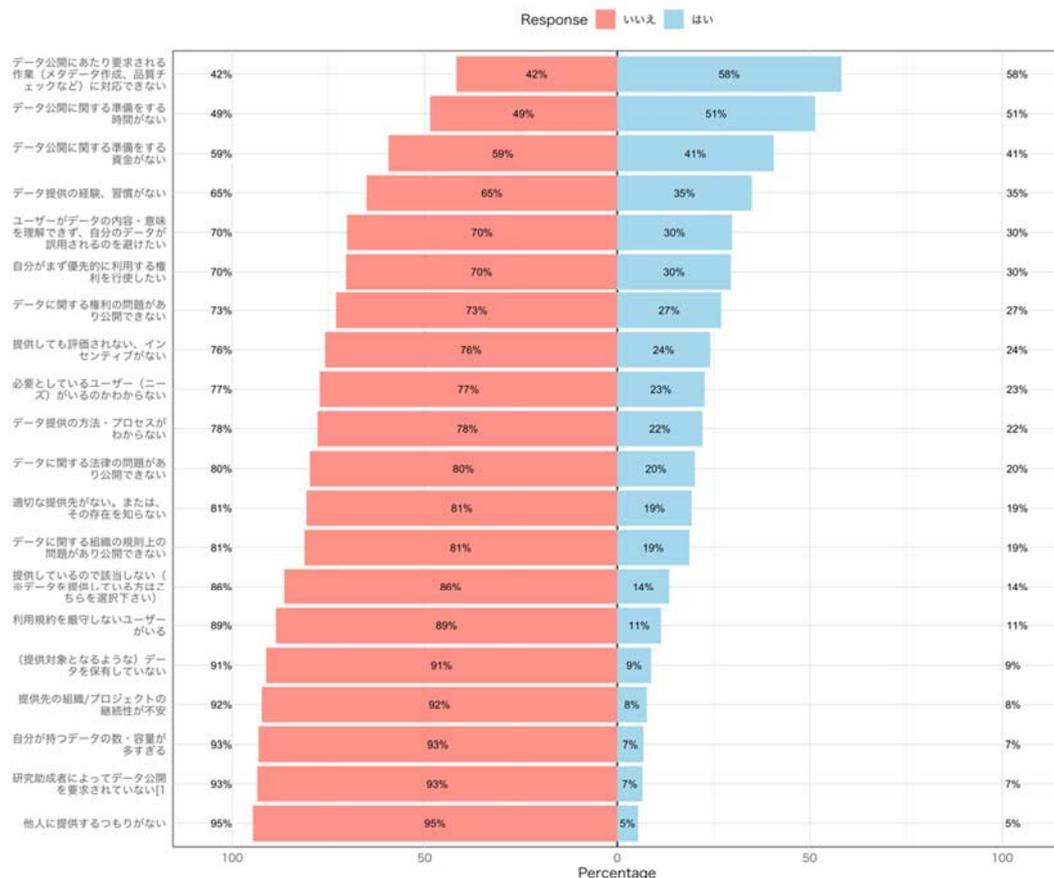


図 16 データ提供できない理由 (Q J14)

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 受けたトレーニング

「データ公開にあたり、今後受けた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 受けたトレーニング (Q K1)

受けたトレーニング	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。
- ・ 第三者による試料・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と読んで、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。そうすると、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いと存じます。

・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ついての知識がなく、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が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%と低いですが、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も多分に含み、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。

公開データを採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る。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」、「適切なりポジトリの選択」、「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、専門家に適切なり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られると思われるため、「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%	15.0%	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	46.8%	23.9%	9.4%	23.0%
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%	8.3%	33.0%
国立社会保険・人口動態研究所	45.5%	45.5%	9.3%	45.5%
国立障害者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%
独立行政法人国立高度医療研究センター	66.7%	33.3%	16.7%	30.0%

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%	23.3%	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%	20.6%	32.4%	23.5%	38.2%	35.3%	50.0%	35.3%	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%
国立社会保険・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%	95.6%	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%	56.0%	33.3%	38.9%	56.0%	61.1%	61.1%	66.7%	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%
独立行政法人国立高度知的学術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入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・利用を求めるといった考え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	割合
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%	48.3%	42.7%
国立感染症研究所	56.7%	25.0%	43.8%	53.1%	45.0%	38.0%
国立研究開発法人医薬基盤・健康・栄養研究所	59.0%	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%	22.5%	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%	62.5%	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%
国立社会保険・人口問題研究所	54.5%	18.2%	63.0%	54.5%	81.8%	36.4%
国立障害者リハビリテーションセンター	68.6%	30.6%	63.0%	77.8%	61.1%	41.7%
独立行政法人労働者健康安全機構	55.0%	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%
独立行政法人国立高度知的学術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ぬこと、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、査読を通じ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う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

- 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.
- Kuipers, T. and J. van der Hoeven, *Insight into digital preservation of research output in Europe*. Survey Report, 2009.
- Likert, R., *A technique for the measurement of attitudes*. Archives of psychology, 1932.
- 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>.
- Tenopir, C., et al., *Data sharing by scientists: practices and perceptions*. PloS one, 2011. **6**(6): p. e21101.
- 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.
- 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.
- Schmitz, C., *LimeSurvey: An open source survey tool*. LimeSurvey Project Hamburg, Germany. URL <http://www.limesurvey.org>, 2012.
- Team, R.C., *R: A language and environment for statistical computing*. 2013.
- Bryer, J., K. Speersneider, and M.J. Bryer, *Package 'likert'*. Analysis and Visualization Likert Items. CRAN, 2016.
- Berners-Lee, T., *Five star open data*. 2009.
- 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.



回答要項 (1)2019年11月XX(金)までにご回答下さいますようお願い申し上げます。
(2)回答には30分程度を要しますので、時間を確保戴くようお願い申し上げます。(3)必須項目は記入しないとアンケートが終了しませんので、Webでの画面表示に従って必須項目の回答終了をお願いします。(4)本研究やアンケート内容等について質問がある場合は木村映善@国立保健医療科学院(kimura.e.aa@niph.go.jp)に直接メールでお問い合わせください。

セクション A: 本アンケートについて

本アンケートは、これまでのオープンデータに関する研究者向けのアンケートをベースに医療分野、公衆衛生分野等の事情を加味したものを追加した内容になっております。具体的には、下記の論文で使用されたアンケートを和訳し、必要な事項を検討・修正・採用したものとっております。

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101> [2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/> [3] Ferguson, Liz. "How and why researchers share data (and why they don't)". WILEY. , (accessed 2016-09-13). <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont> [4]小野雅史, 小池 俊雄, 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 59, 514-525, 2016.
[5]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[調査資料-268]の公表について,
<https://www.nistep.go.jp/archives/35219> (2017).

セクション B: 個人情報の利用について (依頼)

アンケートの進捗管理の為にCookieを利用しておりますが、本アンケートの進捗管理のためのみに利用しています。但し、複数の方が同一PCを利用して回答するようにはなっておりませんので、大変御手数ですが、個別の端末からご回答頂くようお願い申し上げます。

このアンケートに回答することにより、メールアドレスを含む個人情報の提供に同意したことになります。メールアドレスは本アンケートにおける自由回答において意図を確認させて頂きたい等、回答内容についての照会のみ用い、他へ提供されることはありません。またアンケートの回答内容は代表研究者及び分担研究者によって共有されます。アンケートの集計結果はアンケート参加者から希望があれば開示致します。集計結果の一部は論文・報告等で公表する予定です。

本研究やアンケート内容等について質問がある場合は木村映善@国立保健医療科学院(kimura.e.aa@niph.go.jp)に直接メールでお問い合わせください。



セクション C: アンケートで使われる言葉について

「データ」の定義について

今回のアンケートで対象としている「データ」とは基本的にデジタルデータ、あるいはデジタル化されたデータ（紙資料であればスキャンしてPDF化したもの等）です。

物理的な試料（生物由来物、鉱物）等は含まれません。但し、試料から採取された測定、検査結果ですとか、3Dスキャンした構造データ等は対象に含まれます。

「オープンサイエンス」の定義について

オープンサイエンスの定義について統一された見解はありませんが、大まかな傾向としては、下記の様な潮流に対する説明となっています。これまで論文をオープンアクセス化して論文のアクセス機会を増やすことによって最新の知見の入手の障壁を取り除き、学術研究を推進しようという動きがありました。さらに、この理念を推し進め、研究者が保有しているデータも公開できるものは広く公開し、市民、民間、他分野の研究領域の研究者に触れる機会を拡大し、一層の学術活動を推進していくこと、研究活動に関する透明性を確保していくこと、人類の財産としてのデータを保護すること、エビデンスにもとづいた政策の検討を可能にすること、総じて社会運営に公正と平等をもたらすこと、といった多様な理念が語られ期待されています。

オープンサイエンスについての解説はこちらの資料がわかりやすいものですので、ご関心があれば一読頂ければ幸いです。

オープンサイエンス概容 国立情報学研究所 オープンサイエンス基盤センター

<https://rcos.nii.ac.jp/openscience/>

セクション D: アンケートの前に

アンケートの回答中には、ブラウザの"戻る"ボタンは利用せずに、ページ下部の「戻る」ボタンを使うようお願いいたします。

アンケートを中断される際は、右上の「あとで続きをする」をクリックしていただくとアンケートを保存・再開することができます。



セクション E: 最初に貴方の基本情報についてお尋ねいたします。

メールアドレスが一致しません。お手数ですがもう一度ご確認の上ご記入をお願いいたします。確認の為の入力欄にもご記入をお願いいたします。

E1. 所属の機関について選択してください。

- 国立保健医療科学院
- 国立医薬品食品衛生研究所
- 国立感染症研究所
- 国立研究開発法人医薬基盤・健康・栄養研究所
- 国立研究開発法人国立がん研究センター
- 国立研究開発法人国立国際医療研究センター
- 国立研究開発法人国立循環器病研究センター
- 国立研究開発法人国立成育医療研究センター
- 国立研究開発法人国立精神・神経医療研究センター
- 国立研究開発法人国立長寿医療研究センター
- 国立社会保障・人口問題研究所
- 国立障害者リハビリテーションセンター
- 独立行政法人労働政策研究・研修機構
- 独立行政法人労働者健康安全機構
- 独立行政法人国立病院機構
- 独立行政法人国立重度知的障害者総合施設のぞみの園
- その他

その他



E2. 研究者としての経験年数について記載ください。（現職場以前の研究キャリアも含めて下さい。）

- 21年以上
- 16～20年
- 11～15年
- 5～10年
- 5年未満

E3. 貴方の年齢

- 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

60歳以上

E4. あなたの性別

- 男性
- 女性

E5. 貴方のメールアドレス

本アンケートからさらに個別の事情をインタビューするために、御連絡を差し上げる可能性があります。御連絡しても差し支えなければメールアドレスをご記載下さい。

確認の為の入力



E6. アンケートの集計結果・報告書の送付を希望いたしますか？ 希望された場合は、入力されたメールアドレスにPDFデータでお送りしますので、上記項目にてメールアドレスを記載ください。送付は来年度以降の予定です。

Yes

No

セクション F: §公開データの利用状況

F1. 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は
まるものをすべてお選び下さい [5]

[5]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[調査資料268]の公表について,
<https://www.nistep.go.jp/archives/35219> (2017).

サーチエンジン (Googleなど)

データ情報のデータベース (Data Citation Indexなど)

学術機関のリポジトリ・データアーカイブ

出版社のリポジトリ・データアーカイブ

出版社や学術雑誌のサイト(Elsevier,Wileyなど)

論文や学術記事の参考文献

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)

政府・国際機関・出版社などの広報・ニュースレター

ブログや一般的なSNS (Facebook, Twitterなど)

学術系SNS (Mendeley, Research Gateなど)

アラートサービス(RSS等)

メーリングリスト

研究者や同僚に尋ねる・教えて貰う

利用していない

その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。



F2. 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？[5]

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。

[5]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[調査資料268]の公表について, <https://www.nistep.go.jp/archives/35219> (2017).

- 出版社 (Elsevier、Wiley-Blackwell、Springerなど)
- 論文データベース (Pubmed Central, J-Stage, 医中誌)
- 学術データアーカイブ (DDBJ[DNAの塩基配列DB]やICPSR[社会科学データ]など)
- プレプリントサーバ (BioRxiv、engrXiv、PsyArXiv、arXivなど)
- データ共有サービス (figshare, zenodoなど)
- コード共有サービス (GitHubなどオープンソースのリポジトリや、CRAN、CPANなど開発言語のライブラリリポジトリ等が含まれます)
- 学術系SNS (Mendeley, Research Gateなど)
- 厚生労働省や総務省統計局などの政府統計を公開しているWebサイト
- 国際連合やOECDなどの国際機関のWebサイト
- LIS (ルクセンブルク所得研究) などの国際的な研究グループのWebサイト
- 特定研究機関のリポジトリ・データアーカイブ (大学やNASAのリポジトリ等)
- 個人や研究室のWebサイト
- その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

セクション G: §データ保持状況

G1. 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)。[2]

その他がある場合、次の質問にて回答をお願いします。

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

- | | 全く重要
ではない | 重要では
ない | どちらで
もない | 重要であ
る | 非常に重
要である |
|------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 科学の発展に貢献する(新しい研究は既存の知識にもとづいて実施可能) | <input type="checkbox"/> |
| 研究が公的に支援されたものであり、研究結果は公的財産のものであるため | <input type="checkbox"/> |



全く重要ではない 重要ではない どちらでもない 重要である 非常に重要である

既存のデータの再解析ができるようにするため

.....

将来においてデータの検証ができるようにするため

.....

学際的な連係を推進するため

.....

他にはないデータであるため

.....

潜在的な経済的価値があるため

.....

G2. **上記以外のモチベーションがある場合は重要度も含めて記載をお願いします。**

G3. **研究データの保持に関する課題について回答下さい[2]**

その他がある場合、次の質問にて回答をお願いします。

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

全く重要ではない 重要ではない どちらでもない 重要である 非常に重要である

持続可能なハードウェアあるいはソフト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がある

.....

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）

.....

データの起源、信憑性が不確かなものになることで、エビデンスとしての利用価値が失われる可能性がある

.....

データの管理場所を特定できなくなる

.....

データ管理を任せた人・組織が今後の管理において期待に添えなくなる（退職、事業中止等により管理ができなくなる、或いは、放棄する等）

.....

アクセスや利用の制約が維持されなくなる（例えばデータ利用ポリシーやデジタル著作権が将来は尊重されなくなる等）

.....

データ管理のための継続的な予算確保の欠如

.....

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

.....



全く重要ではない 重要ではない どちらでもない 重要である 非常に重要である

自然災害などによるデータ保管場所の損傷

ヒューマンエラーによるデータの損失

G4. 研究データの保持に関する課題について、上記以外の回答がある場合は重要度も含めて記載をお願いします。

G5. 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。[2]

その他がある場合、次の質問にて回答をお願いします。

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

有用ではない やや有用である 有用である 非常に有用である 知らない

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

データ保持に関するワークショップの開催

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構築、フォーラムの開催

ユーザーに向けたデジタルデータ保持に関するトレーニング

G6. 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について、上記以外の回答がある場合は重要度も含めて記載をお願いします。

セクション H: §データ保持の対象

H1. 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。[2]

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

オフィス文書(Word,PowerPoint等)



ネットワーク上のデータ(Webサイト、電子メール、チャットのテキストデータ)

画像(JPEG,GIF,PNG等)

医用画像(DICOM等)

テキスト(テキストエディタ等で編集できるプレーンテキスト)

アーカイブされたデータ(zipファイル等)

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)

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)

ソースコード

Rawデータ(観測機器から得られる生のデータ)

マルチメディアデータ(音声、動画等)(例:手術症例動画、インタビュー録音等)

構造化されたテキスト・データ
(アンケート、調査票、行政記録、医療情報交換文書(CDA)等)

構造化されていないテキスト・データ(インタビュー調査、意見交換のメモなど)

設定データ(ソフトウェア、データベース、機器などの設定ファイル)

構造化されたグラフ(グラフの描画に関する指示やデータが含まれたもの)

その他:具体的な内容について記載下さい。

その他:具体的な内容について記載下さい。

H2. 貴方の研究分野をよく表している分類を最大5つまで記載ください。
。分類については、e-Radにおける研究分野一覧の「分科」にもとづいています。[1] 分科名はe-Radの
<https://www.jst.go.jp/coi/download/file/keyword.pdf>
から採用しています。

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

test

コメント

test

コメント



test



コメント

test



コメント

test



コメント

H3. **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。 [1]**

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

非生物的調査 (土壌、気候、水文学) Abiotic surveys (soils, microclimate, hydrology, etc.)

生物学的調査 Biotic surveys

データ・分析モデルの開発 Data models

実験 (操作・介入を含む) Experimental (involving some degree of manipulation)

インタビュー Interviews

観察 (非介入) Observational (no manipulation involved)

リモートセンシングされた非生物データ Remote-sensed abiotic data (including meteorological data)

リモートセンシングされた生物データ Remote-sensed biotic data

社会科学調査 Social Science Survey (インタビュー調査を含む)

政府統計、国際機関の統計

行政記録・行政資料 (行政記録の例: 個別相談資料、給付記録
行政資料の例: 政策、手続きなど一般に提供される資料)

文献資料 (古記録、新聞、書籍)

その他 Other : もし、他の分野のデータがある場合は記載ください。 

その他 Other : もし、他の分野のデータがある場合は記載ください。



H4. 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。

註釈：今回のアンケートで対象としている「データ」とは基本的にデジタルデータ、あるいはデジタル化されたデータ（紙資料であればスキャンしてPDF化したもの等）です。物理的な試料（生物由来物、鉱物）等は含まれません。但し、試料から採取された測定、検査結果ですとか、3Dスキャンした構造データ等は対象に含まれます。

本回答は今後の国研のデータポリシーの検討にあたって非常に重要な情報となりますので、上記選択肢に当てはまらないものがありましたら、ご面倒でも可及的に記述して頂けますと幸いです。また、このカテゴリに当てはまらないもの、あるいは回答についてご判断に迷われることがありましたら、代表研究者にお問い合わせ下さい。

- 観察データ
- 化学物質（分子）
- 生体・生物由来物
- ゲノム
- 細胞
- 組織・器官
- 生体情報
- 生体反応
- 医用画像（医用波形含）
- 社会経済情報
- 意識
- 行動
- 社会経済的地位、現状
- ライフヒストリー
- 生活習慣・ライフスタイル
- 収入・資産
- 公的支援
- マクロ的情報（国、地域ごとの集計）
- 実験データ
- シミュレーションデータ



その他：もし「その他」データがある場合はご記載下さい。



その他：もし「その他」データがある場合はご記載下さい。

セクション I: §学際的な研究データの利用

11. 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)[2] (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)

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

データの公開はしていないし、公開する予定もない

自分の研究グループのみにデータを開示している

データは全ての人に開示している

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

データへのアクセスは一時的に制限している (例:公開まで1年の猶予期間を設けている等)

データの公開はしていないが、公開を検討するつもりはある

データの公開はしていないが、適切な環境が提供されれば公開できる

自分の所属している学術領域には公開している

有償でデータを公開している

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：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

12. 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)。[2]

[2] Kuipers, T. and van der Hoeven, J. Insight into digital preservation of research output in Europe. PARSE. insight, European Commission, 2009. <https://libereurope.eu/first-insights-into-digital-preservation-of-research-output-in-europe/>

法律等(個人情報保護法など)の制約がある

データの間違った使い方をされる可能性

データの非互換性



技術的基盤がない

予算がない

最先端の研究のアドバンテージを失う（データのフリーライドや論文執筆を先行される等）

データアーカイブへのアクセスが限られている

特に問題はない

その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

セクション J: §研究データの提供状況

J1. 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
:[1]

その他がある場合、次の質問にて回答をお願いします。

(*1) 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。
(*2) Wileyの調査票より組み込んでおります。 (*3) 木村研究班独自項目。

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

	使用して いない	たまに使 用する	時々使用 する	常に使用 している
所属機関のWebサイト・リポジトリで	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
代表研究者のWebサイト・リポジトリで	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
研究者個人のWebサイト・リポジトリで	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
学会や助成機関のWebサイトで (*2)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
一般公開の為のリポジトリ利用(例 Dryad、figshare) (*2)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Journalに掲載されたPaperの補遺として公開	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
研究者間に閉じた環境で(*1)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CDやUSB等の物理媒体で直接 (*3)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



J2. 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。

J3. 貴方が研究データを収集および準備・使用方法についてお尋ねします。 [1]

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

	強く否定する	やや否定する	同意も否定もしない	やや同意する	強く同意する
他の研究者による研究データを収集するプロセスに満足している	<input type="checkbox"/>				
自分のデータを検索するプロセスに満足している	<input type="checkbox"/>				
自分のデータをカタログ化、説明を作成するプロセスに満足している	<input type="checkbox"/>				
プロジェクト期間（短期間）にデータを保存するプロセスに満足している	<input type="checkbox"/>				
プロジェクトの期間（長期間）を越えてデータを保存するプロセスに満足している	<input type="checkbox"/>				
データを分析するプロセスに満足している	<input type="checkbox"/>				
メタデータを準備するツールに満足している	<input type="checkbox"/>				

J4. 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。 [1]

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

	強く反対否定する	やや否定する	同意も否定もしない	やや同意する	強く同意する
私の組織またはプロジェクトには、プロジェクトの存続期間中（短期）にデータを管理するための正式に確立されたプロセスがあるべきだ。	<input type="checkbox"/>				
私の組織またはプロジェクトには、プロジェクト（長期）の存続期間を超えて（長期）データを保存するための正式な確立されたプロセスがあるべきだ。	<input type="checkbox"/>				
私の組織またはプロジェクトは、プロジェクトの期間中（短期）にデータ管理に必要なツールと技術サポートを提供しているべきだ。	<input type="checkbox"/>				
私の組織またはプロジェクトは、プロジェクト（長期）の存続期間を超えて（長期）データ管理に必要なツールと技術サポートを提供しているべきだ。	<input type="checkbox"/>				
私の組織またはプロジェクトは、データ管理のベストプラクティスに関するトレーニングを提供しているべきだ。	<input type="checkbox"/>				



強く反対
否定する

やや否定
する

同意も否
定もしな
い

やや同意
する

強く同意
する

私の組織またはプロジェクトは、研究プロジェクト（短期）の期間中のデータ管理をサポートするために必要な資金を提供しているべきだ。

.....

私の組織またはプロジェクトは、プロジェクト（長期）の存続期間を超えた（長期）データ管理をサポートするために必要な資金を提供しているべきだ。

.....

J5. 貴方が所属している研究分野でのデータ利用に関する課題、懸念事項について意見をお伺いします。 [1][4]

その他がある場合、次の質問にて回答をお願いします。

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101> [4]小野 雅史, 小池 俊雄, 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 59, 514-525, 2016.

強く否定
する

やや否定
する

同意も否
定もしな
い

やや同意
する

強く同意
する

他の研究者や研究機関によって生成されたデータへのアクセスの欠如は、科学の進歩にとって大きな障害である

.....

他の研究者や機関によって生成されたデータにアクセスできないため、科学的な探究をする機会（能力）が制限されている

.....

データの管理（保存・カタログ）が大変である

.....

利用したいデータが分散しており、収集するのが大変である

.....

利用したいデータに関するドキュメントが未整備あるいは内容が不十分である

.....

自分のデータの提供（共有・公開）の進め方がわからない

.....

データの検索、ダウンロードに時間・コストがかかる

.....

データのフォーマットの種類が多すぎる

.....

データフォーマットが統一されていない

.....

データが標準規格に準拠されていない

.....

データが生成された時の採取方法、統計手法がわからない

.....

細かい条件でデータを検索できない

.....

データの複雑さや説明不足により、データが誤って解釈される可能性がある

.....

データの品質が低いことにより、誤った解釈が導出される可能性がある

.....

データ採取当初の意図とは異なった方法で使われる可能性がある

.....



J6. 貴方が所属している研究分野でのデータ利用についての課題、懸念事項について具体的にご記載ください。

J7. データ共有についてお尋ねいたします。[1]

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

	強く否定 する	やや否定 する	同意も否 定もしな い	やや同意 する	強く同意 する
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使用したいと考えている。	<input type="checkbox"/>				
少なくとも一部の自分のデータを制限無しで中央の（公開された）データリポジトリに配置・公開したいと考えている。	<input type="checkbox"/>				
自分の全てのデータを制限無しで中央の（公開された）データリポジトリに配置・公開したいと考えている。	<input type="checkbox"/>				
もしデータへのアクセスに何らかの条件を課せられるならば、私のデータの公開に前向きになれる。	<input type="checkbox"/>				
様々なソースからのデータを統合して研究上の質問に対処できることに満足している。	<input type="checkbox"/>				
様々な方法でデータを使用する幅広い研究者（異なる研究分野の）グループで私のデータを共有したいと考えている。	<input type="checkbox"/>				
他の研究者が私のデータを利用するときに、私のデータが引用されることが重要であると考えている。	<input type="checkbox"/>				
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。	<input type="checkbox"/>				

**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？[5]**

その他がある場合、次の質問にて回答をお願いします。

[5]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[調査資料-268]の公表について,
<https://www.nistep.go.jp/archives/35219> (2017).

	強く反対 する	やや反対 する	同意も反 対もしな い	やや同意 する	強く同意 する
そのままでは難しいが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	<input type="checkbox"/>				
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	<input type="checkbox"/>				
技術的な問題ではなく、単に他分野ではニーズはないと思う	<input type="checkbox"/>				
他分野での利用可能性はわからない	<input type="checkbox"/>				



J9. 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。

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。 [1]

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

はい いいえ

- | | |
|--|---|
| 当該データを使用した論文の共著者としてもらう | <input type="checkbox"/> <input type="checkbox"/> |
| 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（統計法等法律にもとづく承認も含まれます） | <input type="checkbox"/> <input type="checkbox"/> |
| 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引用が必要である | <input type="checkbox"/> <input type="checkbox"/> |
| 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への謝辞が必要である[4] | <input type="checkbox"/> <input type="checkbox"/> |
| プロジェクトで協力する機会（分析に関する相談、結果の解釈、研究成果の普及）を提供してもらうこと | <input type="checkbox"/> <input type="checkbox"/> |
| データ提供者の承認なしに、いかなる形態でも（部分的であっても）データに基づく成果を公表することはできない | <input type="checkbox"/> <input type="checkbox"/> |
| データの採取、取得、あるいは提供に関するコストを部分的にでも負担頂く | <input type="checkbox"/> <input type="checkbox"/> |
| データ利用そのものについて承認の必要はないが、データ提供者が結果のレビューや提案をする機会がない場合は（部分的であれ）提供データに関する結果を公開することはできない | <input type="checkbox"/> <input type="checkbox"/> |
| データ提供者にデータを利用した研究の論文の別刷りを提供すべきである | <input type="checkbox"/> <input type="checkbox"/> |
| データ提供者は利用したデータに関する論文、プレゼンテーション、教材等の全ての成果物に関するリストが与えられること | <input type="checkbox"/> <input type="checkbox"/> |
| データ利用に関する法的許可が必要である | <input type="checkbox"/> <input type="checkbox"/> |
| データの相互共有に関する相互合意が必要である | <input type="checkbox"/> <input type="checkbox"/> |
| 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 | <input type="checkbox"/> <input type="checkbox"/> |



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[4]

CC(Creative Commons)の各条件については、こちらをご覧ください

<https://creativecommons.jp/licenses/>

以下の選択肢は複数組み合わせることが出来ます。例) CC-表示-非営利の場合は 'CC-表示' と 'CC-非営利' の両方にチェックを入れる

[4]小野 雅史, 小池 俊雄, 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 59, 514-525, 2016.

CC-表示(データ作成者についてクレジットを表示することを求める)

CC-継承(元のライセンスと同じ条件下での公開(配布)を認める)

CC-非営利(営利目的の利用を禁止すること)

CC-改変禁止(元のデータを改変しないこと)

犯罪・違反時に利用禁止

犯罪・違反時に罰則

その他:具体的なライセンス形態について記載ください。

その他:具体的なライセンス形態について記載ください。

J12. 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。

J13. 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)。

閲覧数

ダウンロード数

リンク数(他のリポジトリ、Webサイトからのリンク数)

オルトメトリクス(SNSやブログの言及数)

引用数(データを利用した文献数)



その他：具体的な内容について記載下さい。



その他：具体的な内容について記載下さい。

J14. 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。[4]

[4]小野 雅史, 小池 俊雄, 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 59, 514-525, 2016.

提供しているので該当しない(データを提供している方はこちらを選択下さい)

データ公開に関する準備をする時間がない

データ公開に関する準備をする資金がない

データ公開にあたり要求される作業(メタデータ作成、品質チェックなど)に対応できない

自分がまず優先的に利用する権利を行使したい

データに関する権利の問題があり公開できない

データに関する法律の問題があり公開できない

データに関する組織の規則上の問題があり公開できない

提供しても評価されない、インセンティブがない

必要としているユーザー(ニーズ)がいるのかわからない

ユーザーがデータの内容・意味を理解できず、自分のデータが誤用されるのを避けたい

適切な提供先がない。または、その存在を知らない

データ提供の経験、習慣がない

提供先の組織/プロジェクトの継続性が不安

研究助成者によってデータ公開を要求されていない[1]

(提供対象となるような)データを保有していない

利用規約を厳守しないユーザーがいる

データ提供の方法・プロセスがわからない

自分が持つデータの数・容量が多すぎる

他人に提供するつもりがない



その他：ご記載ください。

その他：ご記載ください。

J15. 上記の理由が解決された場合、研究データを公開したいと思われ
ますか？

はい

いいえ

わからない

J16. 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。 [1]

[1] Tenopir, C. et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011 vol. 6, no. 6
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

DMPの提出を求められたことはない

厚生労働科学研究費

日本学術振興会 学術研究助成基金助成金

国立研究開発法人科学技術振興機構(JST)

日本医療研究開発機構 (AMED)

民間助成団体

その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

J17. 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。



セクション K: §データに関する技能

K1. 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[5]。

[5]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[調査資料268]の公表について, <https://www.nistep.go.jp/archives/35219> (2017).

適切なデータ形式

適切なりポジトリ

適切なメタデータ（書誌情報。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・付加情報）

データのバージョン管理法

データのバックアップ方法

データの安全な管理方法（セキュリティ）

匿名加工の基準と実施方法

知的財産権やライセンス

特にトレーニングの必要はない

その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

K2. 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[5]。 データキュレーターとは、資産として残すべきデータを選別したり、データ整形、保存、維持管理等のデータの品質向上、維持管理に寄与することを専門にした職能です。

[5]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[調査資料268]の公表について, <https://www.nistep.go.jp/archives/35219> (2017).

適切なデータ形式への変換

適切なりポジトリの選択

適切なメタデータ標準の選択

メタデータの作成

再利用性があるデータに整える

機関のリポジトリによるデータ公開手続き

データを異分野の研究者に紹介する



匿名加工の実施

知的財産権やライセンス

その他：具体的に記載下さい。

その他：具体的に記載下さい。

K3. 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。

データの引用回数

データのダウンロード数

データ提供者としてのクレジット（論文中の謝辞等含む）

データ採取の従事時間・従事プロジェクト数

データの共有件数（リポジトリへの登録回数）

その他：具体的な内容について記載下さい。

その他：具体的な内容について記載下さい。

K4. 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。

セクション L: アンケートご協力の御礼

これで、本アンケートの質問は全てです。

「次へ」のボタンを押しますと提出が完了し、本アンケートは終了いたします。

見直しの必要がある場合はブラウザの「戻る」ボタンではなく、

本システムの「前へ」ボタンを押してご確認下さい。

ブラウザを閉じて大丈夫です。ご協力いただき、誠にありがとうございました。

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）
分担研究報告書

機微なデータの公開と共有
ANDS ガイドからの機微なデータの公表と共有に関する主要メッセージ

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

国立研究機関所属の研究者のアンケート結果から、機微な個人情報を含むデータは公開することが困難ではないかという意見が散見された。オープンデータと言え、データを一般に広く公開するという印象があるが、最終的な理念はデータの二次利用の機会を最大化することにある。すなわち、データをただちに公開できなくても、そのデータの存在を知らしめることに意義がある。ANDS ガイドラインは機微なデータを公開するにあたっての方法論を紹介している。本稿はそのガイドラインを抄訳し紹介するものである。

- 簡単に適切な手順を踏めば、機微データを公開することの利点は、潜在的な不利益を遙かに上回る可能性があります。
- データを公表する、あるいはデータの記述（つまりメタデータ）だけを公開することによって他の人がそのデータを発見して引用できるようになります。
- データ自体をオープンにアクセスできるように公開せずとも、データの説明のみを公開できます。
- 公開されたデータへのアクセスに条件を付けることができます。
- 非特定化されたセンシティブなデータを公開できます。

A. イントロダクション

A.1 このガイドについて

“ジャーナルや助成機関の方針でデータの共有が義務付けられているにもかかわらず、データをどのように共有すべきかについての実践的なガイダンスはほとんどありません。

んでした...” (Hrynaszkiewicz et al. 2010[1])

本ガイドは、オーストラリアにおけるセンシティブな研究データの公表と共有に関するベスト・プラクティスの概要を示したものである。本ガイドは、「センシティブデータの公開と共有ポスター（第 1.2 節）」で概説されているように、センシティブデータの公開と共有に必要な一連の手順に従っている。このポスターに記載されている手順の詳細と文脈を提供している。興味のある方のために、さらなる読み物のための参考文献が提供されている。以下のセクションとその中のステップに従うことで、データを安全に共有することについて、明確で合法的、倫理的な決定を下すことができるようになります。ほとんどの場合において可能です。

ANDS ガイド¹とセンシティブデータ共有の成功事例の紹介

Sarah Olesen 博士は、センシティブデータとその公開に至るステップを定義している。Leigh Tooth 准教授の女性の健康に関する

¹ 原著の名前は Australian National Data Service

Guide。以下 ANDS ガイドとして言及

る縦断的研究が、どのようにして 20 年間にわたって健康と社会のデータを安全に共有し、大成功を収めてきたかを説明している。

この話は YouTube で視聴できる。
<https://www.youtube.com/watch?v=5j9ftReZc5g&feature=youtu.be> [38mins]。

このガイドとあなたの機関の方針との相互作用

このガイドは、データ管理と公開に関する機関の方針をサポートすることを目的としている。ほとんどの研究者は、所属機関の方針や資金調達の取り決めの中で活動しているため、データの公開に関する決定がこれらの方針に沿ったものであることを確認しなければならない。特に知的財産に関連しており、時にはセンシティブなデータの分類（例：NSW 政府環境・遺産局のセンシティブなデータ種のポリシー）やデータレポジトリの選択にも関連している。

A.2 機微なデータ公開と共有についてのポスター

オリジナルのポスターはこちらで公開されている。

http://www.andis.org.au/_data/assets/pdf_file/0010/385309/sensitive-decision-tree.pdf

翻訳したポスターは別途報告書に同梱している。

「ANDS データ GL オリジナル版」は、上記のオリジナル版を翻訳したものがある。

「ANDS データ GL 国内制度対応版」は、上記のオリジナル版の翻訳をベースに、木村が国内制度との整合性を配慮して加筆修正した案である。下記、国内制度対応版のポスター中の文言についての解説を記す。

「他者によって収集されたデータを使用しているならば、そのデータについて引用してください。」

この文言は、他の研究者・事業者が作成したデータを二次利用して新たなデータが作成された場合、そのデータを元に作成したことを明記することを求めている。これまでのデータの引用については、データを作成しているプロジェクトやWebサイト等の粒度での言及にとどまり、データそのものの引用ではないものが散見されていた。その理由として、当該データを利用している研究者であれば、どこにどのようなデータがあるかの知識は常識的なものであり、また筆者に確認すればいつでも対応できるという状況から、厳密にデータ引用をしていくというモチベーションはなかったようである。ただ、今後は、データ作成者に対して正当な評価を与えられるように定量的な評価指標を収集していく過程において、データ引用の標準化と集計可能な状態で記述をしていくことが不可欠であることから、引用についても今後は可能な限り標準的な手法で行うことを要請したい。

A.3 何故機微なデータを公開、共有するのか？

Box1. 用語の定義

データの公開と共有 (Data publication vs sharing) : 公開 (publication) とは、データが一般に公開される (made public) ことである。これには、一般公開されたデータについての説明と、データ自体へのアクセス、または条件付きアクセスに関する情報が含まれる。データの共有は、データを他の人が利用できるようにする場合をさすが、常にデータの公開を伴うわけではない (例: 同僚間でデータを共有しているが、一般には公開されておらず、外部から発見できない場合や利用できない場合)。

メタデータ (Metadata): データに関する

データ。メタデータの一般的な例としては、図書館のカatalogレコードがある。

リポジトリ (Repository) : メタデータやデータが保管されている場所。データリポジトリには通常、一般の人がデータを検索したり発見したりできるオンラインポータルがある。機関別、分野別、一般的なリポジトリがある。

非特定化データ (de-identified data) : データの人や対象者が特定されるリスクを取り除くために、修正されたデータ。

機微性の高いデータは、データの公開と共有に関する議論からしばしば除外されてきた。機微性の高いデータの共有は倫理的ではない、あるいは安全に行うことが難しくすぎると考えられていた。この見解は、データを「脱機微化' de-sensitise' 」(すなわち非特定化) する方法の理解と利用が進むにつれて変化してきた。出版社や資金提供者が研究者にデータを公表し、共有することを求めるようになったことで、機微性の高いデータの共有も増加している[2 3 4 5]。オーストラリアやその他の国には、機微性の高い情報を含む重要かつ質の高いデータセットが存在する。これには、健康に関する疫学調査、医療試験データ、絶滅危惧種の生態学的研究などが含まれる[6 7]。そして、ほとんどの場合、被験者を特定できるようなデータセットの一部を削除したり修正したり、データへのアクセスや再利用に条件を付けたりすることで、このようなことが可能となる。データを共有することの利点は、出版社や資金提供者の要件を満たすだけではない。研究者、研究機関、研究参加者にとってのメリットは明らかである。

・データやデータセットの説明が公開されれば、他の人が発見することができるので、引用したり、元のデータ収集者や所有者への帰属を表明させたりできる。

・公開されたデータが添付されている学術論文は、それが無い論文よりも引用される頻度が高いというエビデンスがある[8]。

・あなたのデータを他の人と共有して再利用することを選択した場合、新たな共同研究や出版物が生まれる可能性がある。

・データを利用できるようにすることで、高価なもの、稀少なもの、再現性のないものの再分析が可能になる。

・医学研究のデータを共有することで、不必要な重複を減らし、これらの研究から得られたデータを最大限に活用することで、臨床研究の被験者の負担を軽減することができる。

・現在の研究環境や変化する研究環境では利用できないような、安全で継続的な保存が可能でストレージを提供する公開リポジトリにデータを保存できる。

B. センシティブ (機微) データとは？

B.1 センシティブデータの定義

センシティブデータとは、個人、種、物、プロセス、場所を識別するために使用できるデータであり、差別、危害、好ましくない注目を引き出すリスクをもたらすものである。殆どの法律と機関の研究倫理ガバナンス下では、センシティブなデータは、例外を除いて、通常、この形式で共有することはできない。

B.2 人に関するデータ

センシティブな人に関するデータとは、一般的には機微な個人情報をさす。個人情報は、個人またはグループを識別するために使用できるものである[10 11 12]。個人情報は直接個人を特定できる情報 (氏名、生年月日、出生、住所など) と、表1の1つ以上の情報が伴った場合、センシティブなものとなる。

表 1. 1988 年のプライバシー法 (Part II, Division I, Section 6) で定義されたセンシティブ情報の種類

<ul style="list-style-type: none"> ・人種、民族的出自 ・政治的意見 ・政治団体の会員 ・宗教的信念または所属 ・哲学的信念 ・専門家または業界団体の会員 ・労働組合の会員 ・性的指向または慣行 ・犯罪歴 ・健康情報（健康情報の定義については、セクション 6FA を参照） ・遺伝情報 ・生体認証情報

アボリジニとトレス海峡の人々のデータ共有と実践に関する専門的な情報は、以下に記載されている。

- ・ AIATSIS Collection Access and Use Policy
- ・ AIATSIS Guidelines for Ethical Research in Australian

B.3 人間のデータを越えて

センシティブデータには、希少種、絶滅危惧種、商業的に価値のある種の所在を明らかにするデータや、その他の保全活動を含むこともある [13]。合意されたセンシティブな環境・生態学的データの定義の必要性についてはよく説明された資料がある [13][14]。オーストラリア政府の環境省は、センシティブの程度は、脅威の種類、脅威のレベル、脅威に対する脆弱性、情報の種類に照らして考慮する必要があると指摘している。

センシティブな生態学的データの定義が明確になっていないのは、この分野ではデータのセンシティブさが時として過渡的な性質を持つことがあるためである。例えば、

カエルの個体数が少なく、ある場所では減少しているにもかかわらず、別の場所では減少していない場合があるときは、前者ではセンシティブが高く、後者ではセンシティブが低いと考えられる。生物多様性グローバル情報ファシリティ (GBIF) の Chapman and Grafton (2008) [14] は、センシティブなデータを「一般に公開された場合、種や保全活動に「悪影響」を及ぼす」情報と定義している。特に、このデータの公表や共有が、危害のリスクを増大させたり、種に悪影響を及ぼしたりする結果となる可能性につながるような情報を意味する。この GBIF レポートはさらに、データ所有者が、データセットのどの要素がこの種のセンシティブに抵触するかを明確にし、その決定事項をデータセットのメタデータに明確に文書化することを推奨している。

B.4 センシティブな生態学的データの記述

ニューサウスウェールズ州のある地域で、数種の鳥類の個体数と繁殖習性を推定するための調査が半年間にわたって実施された。この調査で得られたデータには、各種の地理的データと繁殖状況が含まれている。今回の調査対象種の一つであるグロッシー・ブラック・オウムは、ニューサウスウェールズ州環境遺産局の「センシティブな種リスト」において「脆弱な」種としてリストアップされている。このリストによると、オウムの脆弱性の理由は「採卵と巣の攪乱の危険性」であるとされている。また、種の「センシティブ」の原因を特定することで、データの所有者はデータセットのどの要素を削除したり修正したりして、データセットの識別を解除できるかを特定できる (Section 5)。

チェック： あなたのデータはセンシティブですか？

あなたのデータに表1の情報、秘密なまたは神聖な慣習 (sacred practice) が含まれている場合、または情報が公開された場

合に種に悪影響を及ぼす結果となるような情報が含まれている場合は、センシティブである可能性がある。データの内容そのものを検討することに加えて、もしデータが共有された場合、そのデータが差別の原因となる可能性があるれば、そのデータはセンシティブなものとなる。後者のこの観点は、データ公開に関する倫理的な考慮事項に関連している（セクション5を参照）。

B.5 文脈に応じた機微性

データの中には、生来的に機微なものもあれば、機微になっていくものもあれば、「機微である」とみなされてしまうものもある。

何が機微であると考えられるかは、時間の経過とともに、また人や被験者のグループ間で異なるかもしれない。機微でないデータは、その文脈によって、あるいはより多くの情報が追加されたときに機微になることがある。組織や出版物の購読者リストの名前は、通常、機微性があるわけではない。しかし、そのグループの特別な関心が購読者を差別にさらす可能性がある場合には、このリストは機微なものとなる可能性がある[12]。例えば、極端あるいは人気がない政治的団体などである。生態学的データの場合、ある種について、ある地域では人間の活動によって被害を受ける危険性があるが、別の地域ではそうではない場合がある。後者の場合、機微性は種そのものではなく、データが参照している場所によって決定されるかもしれない。明らかに機微ではない（すなわち、名前や生年月日が含まれていない）データ、または非特定化されたデータが、文脈が変われば再び機微になる場合もある。2つの一般的な例がある。

1. Triangulation：参加者または被験者の身元または機微な情報が、機微性の低い情報をいくつか組み合わせて特定できる場合。

例えば、人に関するデータでは、年齢、職業、家族構成などの情報があれば、比較的小さなサンプルでも、その人を特定することは難しくない可能性がある。

2. データのリンケージ

データのリンケージとは、同一人物や研究対象を含む2つ以上のデータセットを組み合わせることである。データセット単体だけでは、個人を特定したり、被験者を危険にさらしたりするのに十分な情報を含んでいないかもしれないが、2つ以上のデータセットと組み合わせることで、個人を特定可能になるかもしれない。例えば、データセットAは、認知障害のある、特定できない患者のグループの病歴を記述している。データセットBは、同じく中程度に小さな人口での雇用情報と公共交通機関の利用状況を含む。これらのデータセットがリンクされると、患者がどこで働き、どこに住んでいるかについての十分な情報があり、患者が特定される可能性がある。生態学研究からの例をあげる。データセットAは、採卵に弱いハヤブサの種と、その個体数と場所を経時的に記述している。データセットBは、同じハヤブサの種の同じ期間の繁殖パターンを含む。これらのデータセットがリンクされた場合、営巣時のハヤブサの位置を決定するのに十分な情報があるかもしれない。データの所有者や管理者は、常にデータセットの中でTriangulationの可能性を検討し、その可能性があるかを確認する必要がある。新しいデータが導入されたときに、データのリンケージが行われる可能性をふまえてTriangulationを再考するのは良い習慣である。

C. 私は機微なデータを持っています - さて、どうしますか？

C.1 機微なデータの公開と共有方法

前のセクションでは、元の形式の機微データについて説明した。ANDSの「非特定化のためのガイド」では、元のデータを修正

して機微性が残らないようにする方法を説明している。この種の修正されたデータは、研究参加者や被験者を危険にさらすことなく、他の研究者にも再利用可能であり、非常に価値のあるものである。また、そのデータは、より少ない法的・倫理的制約の下で公開・共有できる。ほとんどの場合、非特定化された機微なデータは条件付きアクセスで公開される。データに関する記述（すなわち「メタデータ」）がデータリポジトリで公開されるので、制限なくデータの存在が発見される状態になる。そして、データ自体は、ある条件（研究者やリポジトリによって設定された）が満たされるとアクセス可能になる。公開されたデータの説明には、アクセスに関する情報と、アクセスを申請する方法や場所へのリンクが含まれている。

C.2 何が合法か？

オーストラリアのプライバシー法（Australian Privacy Act 1988[10]）の下では、機微性の高いヒトや個人のデータは一般的に元の形のままで共有できない。しかし、いったん非特定化された後は、これらの修正されたデータはもはや同法の対象とはならない。言い換えれば、非特定化された機微なデータは合法的に共有することができる。1988年オーストラリア・プライバシー法では、非特定化の定義を「個人情報、情報が識別可能な個人または合理的に識別可能な個人に関するものでなくなった場合に非特定化される」としている（第2部第1部第6節）。プライバシー法はまた、「識別情報」の定義を提供しており、個人をデータから識別特定できないようにするためには、最低限どのような情報を削除しなければならないかを示している。プライバシー法（1988年）は、識別情報が除去されたデータには適用されないが、データを非特定化する行為（すなわち、元の機微性の高いデータセットから識別情報を除去する行為）には適用されることに注目すべき

である。しかし、この行為は、プライバシー法（1988年）のオーストラリアのプライバシー原則の中で、機微性の高いデータの利用に対する数少ない例外の一つとして明示的に容認されている。これは、特定の同意を必要とせずに「個人が自分の個人情報を使用されたり開示されたりすることを合理的に期待することができる」「通常の...慣習」と考えられているからである[16]。データ共有における同意の詳細については、「倫理的配慮」のセクション（セクション5）を参照すること。

機微なデータの共有の合法性に関する重要なポイント

- 機微性の高いデータは、ほとんどすべての場合、元の形で公開したり共有したりすることはできない。
- 非特定化されたデータは、もはや機微なものではなく、共有することができる。
- 人や個人のデータを非特定化する行為は、オーストラリアプライバシー法（Australian Privacy Act 1988）の下で認められている[16]。

D. データの非特定化

データ共有のための非特定化の目的は、研究の参加者や被験者（例えば、動物や植物の種）が識別され、危害や差別の危険にさらされることを防ぐことである。これには、元の（機微性の高い）データセットの情報を削除したり修正したりすることが含まれる。削除したり修正したりする必要がある情報は、データセットの内容や、データが機微であるとみなされた理由によって異なるだろう。

チェックポイント：私のデータを非特定化することはできますか？

データの中には、参加者や被験者への危

害のリスクを十分に最小化するために修正できないものや、修正した場合でもそのデータの利用価値が損なわれてしまうものがある。これは、データが収集された文脈を配慮することで身元の特定可能になることが多いからである。例えば、踏み込んだ個人面接からのデータ、および アイデンティティがデータの本質的な側面である文化的または歴史的なデータなどである。前者の場合、データは研究に参加していない（したがって同意していない）他の個人の アイデンティティも明らかにすることがある。大多数において、参加者からの明示的な同意およびヒト研究倫理委員会の承認なしには、非特定化できないデータを公表することはできない。ただ、メタデータが研究に参加した個人または被験者を特定するために使用できない場合には、必ずしもメタデータの公開を妨げるものではない。

D.1 生態データ

環境データや生物多様性データの手法は、ヒトデータに比べて確立されていない。ヒトデータの場合、機微性は通常、個人を特定することに関係している。したがって、非特定化するためには、識別情報を除去しなければならない。しかし、生態学的データの機微性は、一般的には種の名前そのものではなく、その場所や繁殖・結実・移動の時期に関する情報に関係している。例えば、第2.4節の例示的なデータセットは、単にグロッキー・ブラックコ・ガトウのデータが含まれているからといって機微性が高いわけではなく、繁殖時の正確な位置に関するデータが付随しているからである。機微性が低下した生態学的データの実例は、オーストラリア政府環境省の「センシティブな生態学的データ - アクセスと管理に関する方針 V1.0」[15] の表1にある。

D.2 再同定のリスク管理

データが他のデータとリンクされている場合、データが非特定化された後に、再特定のリスクを見直さなければならない。データリンケージとは、同じ人や被験者のデータを含む 2 つ以上の別々のデータセットをマージすることである。疫学、医学、社会科学、生態学の分野ではますます一般的になってきているが、その理由は、研究者が新たなデータを収集しなくても、より多くの情報を追加することで、人々や被験者の文脈をより詳細に理解できるからである。また、データリンケージは、既存のデータセットからより大きな価値を引き出すことができる。Triangulation と同様に、データリンケージは、(新たにリンクされたデータによって) 1 つまたは複数の潜在的に識別可能な情報が追加されたために、識別されていなかった参加者や被験者が再特定される可能性があることを意味している。データがリンクされる際には、新しいリンクされたデータセットを(潜在的に) 識別可能なデータセットとして扱い、そのリスクを評価することで、この再特定可能性を評価しなければならない。

D.3 いまのうちに

公開のためにリポジトリに提出するデータセットは、以下のようにしなければならない[1]。

- ・クリーンアップ; エラー、外れ値、重複、欠落データをチェックする (本当に欠落しているデータにはそのような注釈をつける)
- ・注釈をつける; 変数とオブジェクト (関連する場合はそれらのカテゴリ) には明確なラベルをつけて記述し、必要に応じてキーを提供すべきである。変数やオブジェクトが非特定化の間に元の形から変更された場合は、そのことをメモしておく。
- ・仕様がオープンであり、変換やアーカイブが容易なフォーマットであること。一般的に使用されている様々なオペレーティ

ングシステムやプログラムで読みやすく、操作しやすいデータ形式を使用するように努める。アクセシビリティを高めるために、非専有（「オープン」）フォーマットも推奨する。データ形式に関する情報は、メタデータレコードの中で提供することが望ましい。データを読み込んで分析するために特殊なプログラムが必要な場合は、可能であればデータと一緒に提供すべきである。データフォーマットに関する特定の要件については、選択したデータリポジトリを確認する。

E. 倫理的配慮

法律や地域社会の基準を満たすことに加えて、研究者は参加者や研究対象者に対して倫理的な義務を負っている。これには、プライバシーを保護し、研究への参加とその後の発表から生じる可能性のある危害を回避することが含まれる。参加者の信頼と研究の完全性を維持するためには、データの倫理的な管理が研究者の第一の関心事でなければならない。

E.1 新規ヒトデータの場合、または研究参加者との接触が可能な場合

データを取得して公開する前に、ヒトの参加者の同意が必要である。同意を得るための最良のタイミングは、データが収集される前である。これは最良の実践であるだけでなく、研究プロセスの後半で同意を得ようとすることで、費用、遅延、データ利用の損失を回避できる。データの共有が要求された場合、参加者が研究への参加を拒否するだろうという懸念は、根拠のないものである可能性が高い。[17 18]

研究参加者に非特定化データの公開と共有についての同意を求めるにあたり、含めべき事項は、

1. 研究への参加に同意する前、およびデータ収集への同意を求められる前に、参加者に提供される「情報シート」に、データの非特定化、公開、共有のプロセスに関する情報を含める。この情報は、参加者にとって理解しやすく、提供したデータの公開および／または共有に同意することについて十分な情報に基づいた決定を下すことができるように、十分な詳細が記載されていなければならない。参加者情報シートは、研究のあらゆる側面が開始される前に、ヒト研究倫理委員会（HREC）の承認を得なければならない。それには、機微保持の手順、データの公表、データ共有が可能な条件（研究を実施している者以外の研究者とデータを共有する見込みがあるかどうかを含む）を（簡潔に）記載する必要がある[6]。

2. データの共有および／または公表に対する同意を「同意書」内で具体的に要求する。同意書においてデータの公開と共有を依頼する際の文言例を以下に示す。

3. データの公表と共有に関して、管轄区域内の倫理申請書、情報シート、同意書の要件を熟知しておく。例えば、National Health and Medical Research Councilのヒト研究倫理申請書（HREA）では、データの共有に関連したいくつかの質問、特に Q3.13～Q3.17が挙げられている。

E.2 データの公開・共有を求める同意書

参加者の同意書におけるデータの公開および共有に関する記述は、以下の通りである。

1. データの非特定化、公開、共有を妨げることを避ける[19]。
2. 将来のデータ公開（リポジトリへの保存を含む）と共有の可能性を述べる。
3. データへのアクセスが他者に認められる条件を記述する。これには、匿名化するプロセスや、元の研究チームの承認などの他

の条件が含まれる可能性がある（セクション G.1「条件付きアクセス」を参照）。場合によっては、参加者が自分のデータを誰と共有することに同意するか（誰と共有しないことに同意するか）を選択する機会を提供することも適切な場合がある。

4. 収集したデータを文書化し、後のデータ利用者が、参加者が同意した条件を認識できるようにする。

E.3 データの公開と共有を求める同意書の例文

データが公開されることを意図している場合、または少しの制限のみでアクセスできる場合：’

この研究の情報は、あなたが誰であることを明らかにしない方法をとおしてのみ使用されます。この研究の出版物や他の研究者と共有されるデータファイルでは、あなたの身元が特定されることはありません。貴方がこの研究に参加していることは秘密にされます。[20]

ファイルが他の研究者と共有されたり、結果が公開されたりする前に、あなたを特定できる個人情報は削除されたり変更されたりします。[18]

私は、研究のために収集された研究データが、私の名前やその他の識別情報が使用されない限り、公表される可能性があることに同意します。

条件付きアクセスが可能なデータの場合：

他の真正な研究者は、このフォームで要求された情報の機微性を保持することに同意した場合に限り、このデータにアクセスすることができます。[21]

上記の例は、その再利用時に適用される特

定のアクセス条件を含むように適応させることができる。

他の真正な研究者が将来的に非特定化されたデータへのアクセスを要求する可能性があります。このフォームで要求された情報の機微性を保持することに同意した場合にのみアクセスが許可されます。また、そのような研究者のアクセスには、元の研究チームの承認も必要となります。

また、参加者に、データ共有に同意（あるいは反対）する人を選択する機会を与えることも検討してください。例えば、データを再利用する可能性の高い人のリストから選ぶことができます。

E.4 既存データの場合、研究参加者との再接触ができない場合

センシティブなデータは、以下の場合には、研究参加者の明示的な同意なしに合法的に共有することができる。1. データ収集の同意に先立って参加者に提供された情報が、データの将来の使用を示唆している場合*、または
2. 同意を得る機会がもはや存在しないか、または現実的ではない場合において、データは非特定化が可能であり、かつ非特定化のプロセスがプライバシー法（1988年）に規定された定義に合致している場合、かつデータの公表または共有が研究参加者または被験者に害を与えたり差別を助長したりする危険性がない場合、最初のデータ収集時の情報シートおよび同意書が共有を排除していない場合。

注：* 参加者の同意書がデータの公開や共有について具体的に言及しておらず（いずれも除外していないこと）、情報シートが言及している場合、プロジェクトへの参加自体に同意があれば、共有が可能である。これは、参加への同意が情報シートへの理解

と同意を意味するからである。

推奨: これらの要件を満たしているかわからない場合は、研究プロジェクト以外の同僚や最寄りのヒト研究倫理委員会に意見を求めることを勧める。

E.5 自分が収集していない機微なデータの共有

あなたがデータの所有者ではない場合、データの公開や共有についてはどうすればいいのでしょうか?つまり、あなたはデータの再利用者または「二次データ利用者」である。データの再利用は、疫学のような非特定化データを扱う研究分野では、すでに比較的広まっている。上記の質問は、学術論文と一緒にデータを公開することを求める雑誌の要求に応えるために、ますます一般的になってきている。ほとんどの場合、自分が収集していないデータは、そのデータの著作権を所有していないため、公開できない。例外は、データ所有者が再配布を許可するためにライセンスを与えたデータの場合である。

研究倫理に関する注意事項:元の研究者と同様に、二次データの利用者にも「データが責任を持って敬意を持って使用され、参加者のプライバシーが守られていることを保証する義務」がある[6]。したがって、データの再利用者は、元の研究者が参加者に指定したデータ利用条件、および再利用のためのアクセスが許可された際にデータ所有者または管理者が提示した再利用条件を守らなければならない。

F. データの引用

もしあなたがデータの再利用者であるならば、そのデータに基づいたすべての論文、プレゼンテーション、助成金申請の際に、そのデータの元の出典を参照したり、引用

したりしなければならない。これにより、データの所有者はデータの利用状況を追跡できるだけでなく、あなたの研究をこのデータセットに基づいた他の論文とリンクさせることができる。この方法の詳細については、ANDS Guide to Data Citationを参照すること。

推奨:

- 再利用するデータがライセンスされている場合は、データ共有に関するライセンスの条件に従うこと。(一般的には「再配布(‘redistribution’)」と呼ばれている)再配布が許可されている場合は、データを共有し、データ所有者に帰属させることができる。あるいは、他の人がアクセスを要求できるように、公開された研究の説明にデータの出所を引用することもできる。
- 再利用するデータがライセンスされていない場合は、データの所有者または管理者に連絡して、公開と共有に関する指示を求めること。

データを再利用する場合には、すべての学術成果物に元のデータの出典を引用する必要がある。

G. データを発見可能にする

リポジトリ内の公開メタデータ(public metadata)は、あなたのデータを他の人が発見できるようにするための最良かつ最も簡単な方法である。以下は、公開カタログでのデータに関する説明である。

「データ自体を自由に利用できるようにしないとしても、データの説明を公開することはできます。」

リポジトリからは公開するにあたり、どのようなメタデータを用意する必要があるかを説明される。ほとんどの場合、助成金

申請書やプロジェクト報告書、論文などで既に持っている情報である。公開されたメタデータは、他の人があなたのデータを見つけて引用することを可能にする。データ自体がメタデータと一緒に公開されていない場合（ただ、多くはメタデータ内のリンクを介して公開されている。以下の例を参照）、メタデータはデータへのアクセス方法に関する情報を再利用者に提供する。

機微性の高いデータを持つ研究者は、自分のデータを公開して共有したいと考えているが、公開やアクセスについてはまだ懸念を持っているのが一般的である。これは、条件付きアクセスでデータを公開することによって、大部分が対応可能になる。

G.1 データへの条件付きアクセス。条件付きアクセスとは何ですか？

条件付きアクセス(conditional access)は、メタデータが公開されている（すなわち公開リポジトリでメタデータが公開されている）が、データ自体へのアクセスは、あらかじめ定められた条件が満たされた場合にのみ発生する。これらの条件は研究者やデータ所有者、および/またはデータリポジトリによって設定される。これらの条件には、データの再利用者となる可能性のある人に以下のことを要求することが含まれる。

- ・ 登録および/または連絡先の詳細を提供すること
- ・ データをどのように使用、保存、または管理するかについての情報を提供する。
- ・ データセキュリティ、プライバシーの条件に同意すること。
- ・ 共同研究やその他の目的でデータ所有者から連絡を受けることに同意すること
- ・ アクセス料金を支払うこと
- ・ 元の研究参加者（ヒト）が同意した同意フォームおよび情報シートに記載されているその他の条件を満たしていること。

機微なデータを発見可能な状態にしたいが、条件付きでアクセスできるようにしたいと考える理由には、以下のようなものがある。

- ・ 再利用者が真正な研究者であることを確認するため
- ・ 再利用者がデータの機微性または安全な保管を維持することを認識し、同意していることを確認するため
- ・ 参加者同意書または研究倫理申請書で条件付きアクセスを指定した場合
- ・ 誰が何のためにデータを使用しているのか、あるいは何のためにデータを使用しているのか、ある程度の監視を維持したいとき

その他の条件付きアクセスの理由としては、以下のようなものがある。

- ・ データ（またはデータの一部）が禁輸下にある場合
- ・ コラボレーションができるように、誰がデータを利用しているのかを知りたいとき

他の人は、データを新たな価値ある目的のために利用できる。研究データは、データ作成者や管理者が（他者による利用に）価値があると考えるかに関わらず、広く普及させるのが理想的である。データの用途が将来的にどのようなものが存在するかを知ることは不可能である。

G.2 データへの条件付きアクセスが可能な公開メタデータの例。オーストラリア女性健康縦断研究(ALSWH)

「女性の健康政策と実践の発展を支援するためのエビデンスベースを強化するための国の研究資源としての ALSWH にとって、デ

ータの共有は基本的なものです。私たちは、私たちのデータを利用できるようにし、幅広い分野の研究者間のコラボレーションを奨励することに全力を尽くしています。学際的なグループ間でのデータ共有は、新鮮な視点と女性の健康に関する新たな洞察と知識を得るための機会を提供します。」

教授 Gita Mishra, ALSWH ディレクター

ALSWH のメタデータレコード

オーストラリア・データ・アーカイブ（データ・リポジトリ）に保管されている女性の健康に関するオーストラリア縦断的研究のデータの「利用条件」は、メタデータ・レコードで明確に説明されている。

データの利用について正式な要請があった場合には、共同研究者がデータを利用できるようにすることができる。データ利用の許可を得る必要がある。

G.3 ライセンス

再利用を目的としたオーストラリアのデータはすべてライセンスが必要である。これには、非特定化されたデータセットも含まれる。ライセンスとは、データの使用方法を明確に定めた文書であり、元のデータ所有者に帰属する。ライセンスがないと、データをどのように再利用できるかが不明確になり、再利用する人の意欲をそぐことになりかねない。

ライセンスには様々な形態があり、再利用に関する制限が少ないものから多くの制限があるものまで様々である。データリポジトリの中には、独自のライセンス文書を持っているところもある（例：Australian Data Archive の「Access Categories」）。また、オープンアクセス、つまり無制限のアクセスを要求するものもある（例：Dryad）。

クリエイティブ・コモンズでは、再利用の条件が少ないものから多いものまで、さ

まざまなライセンスを提供している。クリエイティブ・コモンズのウェブサイトでは、簡単な手順でライセンスを適用することができる。

注意すべき点は、あなたのデータにオープンライセンスを適用したとしても、機微性の高いデータを公開したり、データの非特定化に代わるものとして機能したりするわけではないということである。機微なデータは、ライセンスがあっても機微なままであるため、参加者の同意なしに公表することはできない。したがって、本ガイドでは、データの非特定化がなされた後、または特定可能なデータを共有することに同意されており条件付きアクセスの取り決めによって利用可能になっている場合のみに、ライセンスを適用し、公表することを推奨する。詳細については、データ再利用のためのライセンスおよび ANDS Research のデータ著作権およびライセンスに関する FAQ を参照すること。

H. 公開する権利

ほとんどの場合、データを公開する個人または機関は、データを公開するための適切な権利（著作権など）を保有していなければならない。ライセンスは、そのデータの権利者のみがデータに適用できる。

H.1 データの著作権が自分にあるかを知るには？

著作権の所有権と著作権放棄の例は、オーストラリアの教育機関によって異なる。著作権は知的財産の一側面であるため、所属機関または雇用主の知的財産ポリシーを調べることを勧める。それでも不明な点がある場合は、所属する研究室または研究サービス部門にアドバイスを求めること。より詳細な情報については、ANDS Guide to Copyright, Data and Licensing を参照すること。

H.2 自分のデータはどこで公開すればよいですか？

多くのデータリポジトリから選択することができる。リポジトリの中には、メタデータのカタログのみを提供し、データの保存場所にリンクしたり、参照したりするものもある。また、メタデータをカタログ化し、データ自体を保存しているリポジトリもある（例：Australian Data Archive、Figshare）。リポジトリには、機関固有のもの、分野やコンテンツ固有のもの、一般的なものなどがある。国際的なリポジトリのリストはre3data.orgで見ることができる。

機微データを公開する場所を選ぶ際には、以下の点を考慮すること。

- ・ リポジトリがデータへの条件付きアクセスをどのように管理しているか、また、条件付きアクセスはリポジトリが管理しているのかデータ所有者が管理しているのか。
- ・ 機関、雇用主、資金提供者、出版社が特定のリポジトリの利用を義務付けているか、推奨しているか。

データをどこで公開するかについてのその他の考慮事項としては、以下のようなものがある。

- ・ あなたの研究分野では、データを公表する場所に関する慣習があるか
 - ・ メタデータとデータを同じ場所で公開するかどうか
 - ・ リポジトリが必要とするメタデータ
 - ・ リポジトリがデータを必要とするフォーマット
 - ・ 発生した経済的コスト
 - ・ リポジトリがデータに一意の識別子を割り当てることでデータ引用の追跡を可能にするか（DOI 識別子等）

I. 謝辞

著者は、本ガイドの執筆にあたり、グレッグ・ラフリン博士（ANDS 主任政策顧問）、バーデン・アプリアード博士（AusGOAL 国家プログラム・ディレクター）、マイケル・マーティン教授（オーストラリア国立大学財務・アクチュアリー研究・応用統計学 研究科、人文社会科学委任倫理研究委員会委員長、オーストラリア国立大学科学・医学委任倫理研究委員会委員長）、ジェフ・トランター博士（Environmental Resources Information Network、ERIN）に貴重な助言とコメントをいただいたことに謝意と感謝を表したいと思います。

J. Reference

原著を参照のこと

この文書は、Creative Commons Attribution 3.0 Australia License にもとづいて Australian National Data Service が公開した” Publishing and sharing sensitive data” を抄訳したものです。元ライセンスにもとづき、Creative Commons Attribution 4.0 International License にて公開いたします。

オリジナルの文書は、こちらからアクセスできます。

https://www.ands.org.au/_data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf

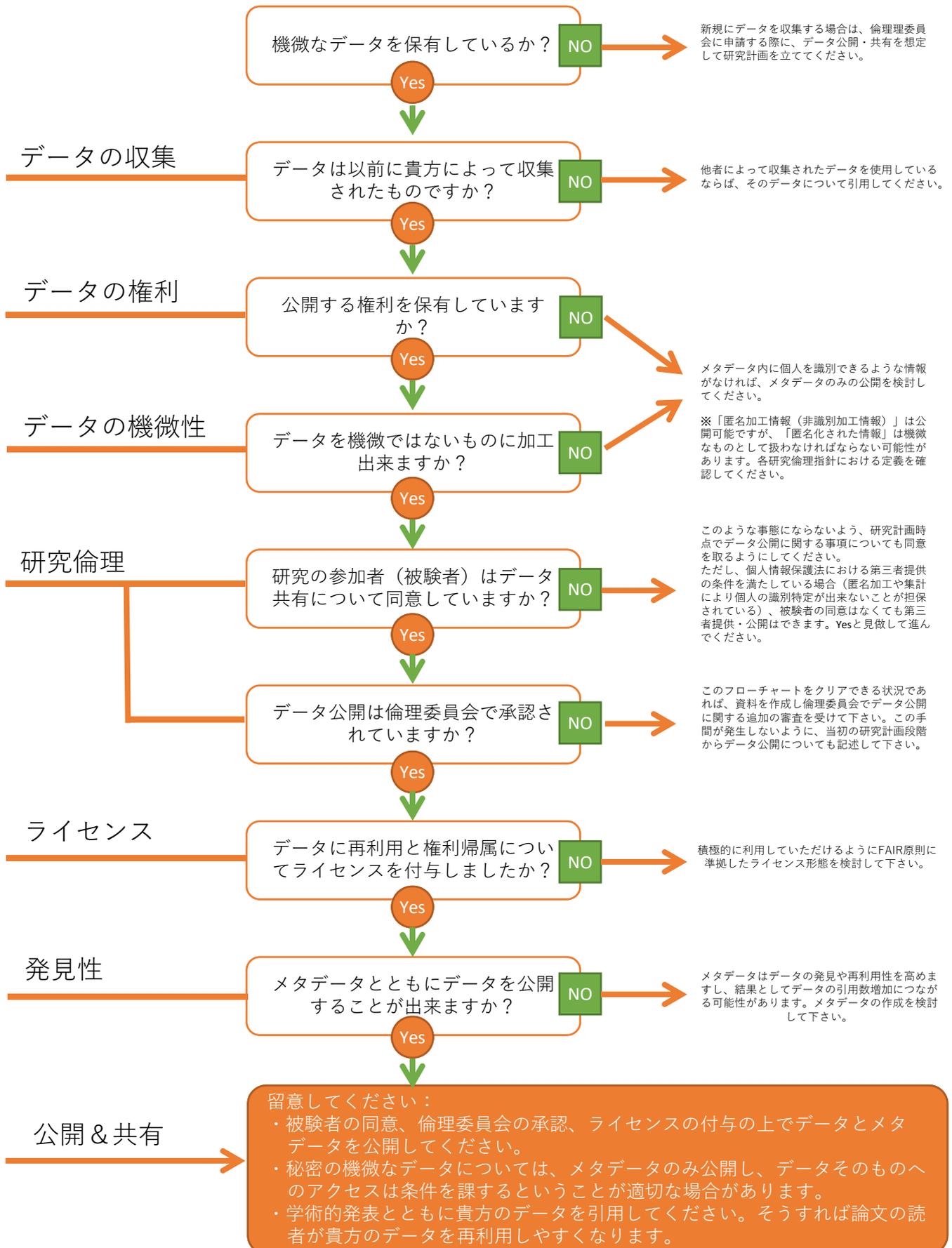


このガイドラインを抄訳した理由は、「機微なデータであるから、オープンデータに供することはできない」という誤解を解くために、このガイドラインが有用であると判断したためです。直接公開できないものに

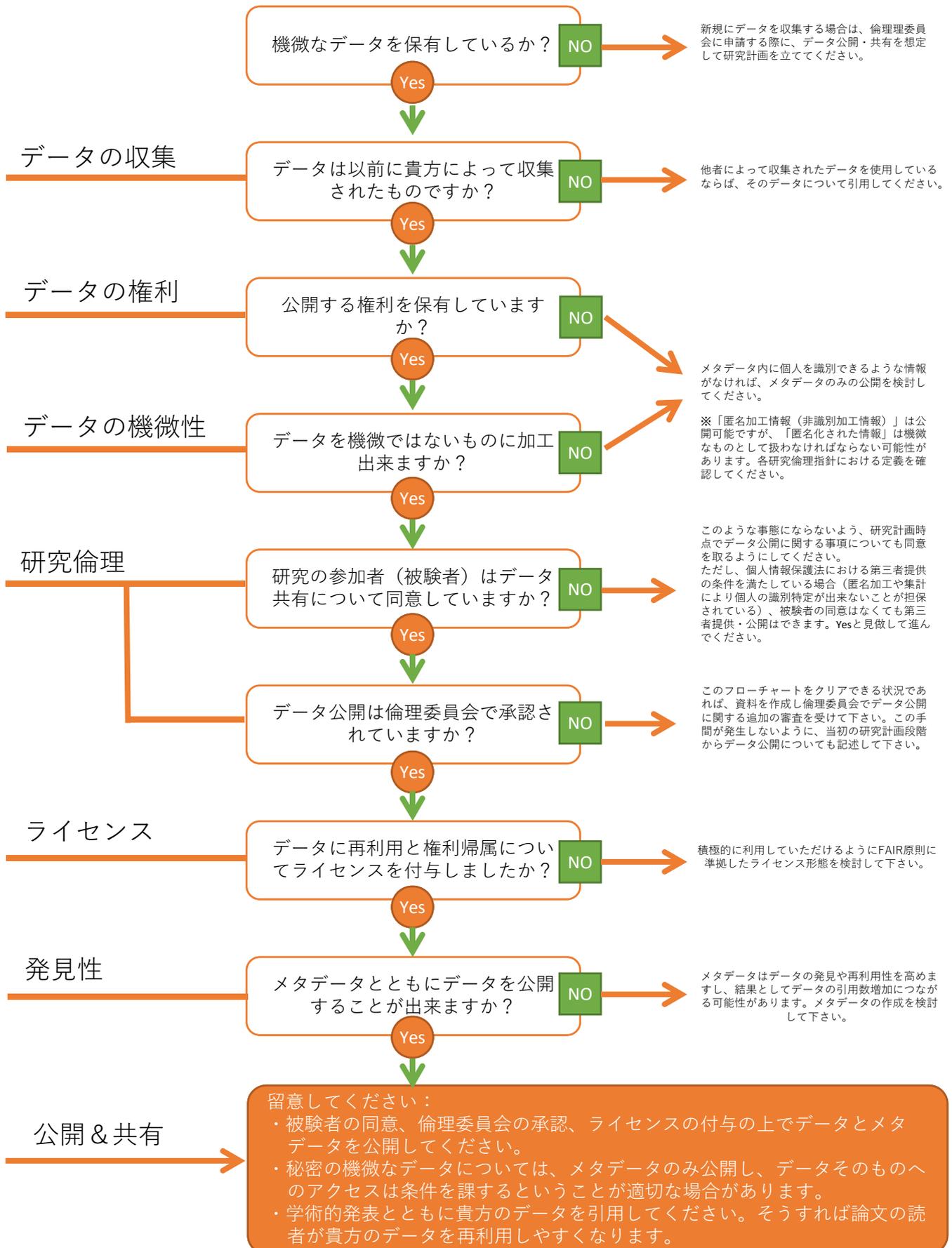
については、データの内容について説明するメタデータのみを公開し、データの発見性、引用、コラボレーションの可能性を向上させることが期待されます。

なお、元文書の趣旨を把握することに要点を置いているため、レイアウト等の厳密な再現は意図しておらず、図面の一部を掲載していません。引用等の際は原著を必ずご確認ください。

機微なデータの発表と共有にむけての判断フローチャート



機微なデータの発表と共有にむけての判断フローチャート (国内制度対応版)



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）
分担研究報告書

国立研究開発法人におけるデータポリシー策定にあたり検討すべき事項の整理
～「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」を羅針盤として～

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

本文書は国立保健医療科学院のワーキンググループにおいてデータポリシーを策定するにあたって作成した参考資料である。ワーキンググループにおいて「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」を参考にしたときに、当ガイドラインのみでは判断が困難な構成員向けに情報提供をすべく調査した事項をまとめたものを、公開向けに再構成したものである。

A. 本文書の位置付け

本文書は国立保健医療科学院のワーキンググループにおいてデータポリシーを策定するにあたって作成した参考資料である。すなわち、ワーキンググループにおいて「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」を参考にしたときに、当ガイドラインのみでは判断が困難な構成員向けに情報提供をすべく調査した事項をまとめたものである。他の研究組織におけるデータポリシーの検討の際に参考になると判断し、公開するものである。なお、この文書中における提言はワーキンググループへの問題提起として記したものであって、国立保健医療科学院の最終的結論ではないことはご留意頂きたい。

B. ガイドライン検討にあたって背景

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」（以下ガイドライン）（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、厚生労働省所管の研究機関（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）だけでなく、その他の研究機関（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）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。

データポリシー策定にあたって考慮すべきことからはきわめて広範囲にわたるので、「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」（以下、DPG: Data Policy Guideline）[1]及び当該ガイドラインの解説資料 平成 31 年 4 月 2 日版（以下、AG: Appendix of Guideline）[2]をデータポリシー策定プロセスの羅針盤とし、逐条的なガイドラインの検討をとおして課題を整理し、2020 年度中のデータポリシー策定にむけて準備すべきことを洗い出す。

なお、AG において、各法人の研究分野の特性やミッションの反映を阻害することが懸念されるため、データポリシーのひな型は敢えて提供されておらず、ガイドラインに沿って各法人にて検討すべきことが記載されている。しかしながら、各法人で委細にデータポリシーを検討できる人員は限

られているため、本稿において、DPG、AG に加えて独自に調査、分析した事項を提供することによって、各法人での検討の一助になることを期待するものである。

以下、内閣府の「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」（DPG）に記載されている内容を元に、逐条的に当院におけるデータポリシーを検討するにあたっての考察を記述する。DPG の「3. データポリシーで定めるべき項目」以降に提示されている項目（番号）と解説を囲みで引用し、その後我々が検討したことを記述する。

C. DPG の逐条的検討

C.1 機関におけるポリシー策定の目的について

・機関のビジョン、ミッション等を踏まえ、ポリシーを策定した背景と研究 データ利活用の目的について記述する。

・機関が Web サイト等で公開している機関のビジョン、ミッションをベースに解説する。

・公的資金を活用して実施した事業・研究を通して得られた成果を国民に還元し、広く利活用を促進することを通し、機関に求められている役割の強化とアウトリーチを拡大することを説明する。

C.2 管理する研究データの定義、制限事項について

・機関のミッションに従い、ポリシーが対象とする「研究データ」の定義・ 範囲を明確にし、利活用が想定されるデータ、将来的に利用の可能性が考えられるデータなど、研究データの種別・内容等について記述する。

・研究データの利活用に関する機関の方針や基本的な考え方を踏まえ、また、第 5 期科学技術基本計画が示すオープンサイエンスの推進に係る方針にも留意し

て、非公開、共有等の対象となる研究データや公開・共有における 制限事項について記述する。

C.2.1 研究データ利活用に関する機関の方針・基本的な考え方

第5期科学技術基本計画 が示すオープンサイエンスの推進に係る方針¹に記述されている、オープンサイエンスの推進体制の基本的姿勢は、公的資金による研究成果の利活用の機会を可能な限り拡大することとしている。我々の活動は、国の予算と公的機関からの競争的研究資金に基づいている。それを踏まえて、基本的に保有しているデータは公開が前提であるとし、非公開となるものについて例外的に指定・除外していく手法を採用することとする。

C.2.2 検討対象となるデータの種類の種類

保健医療科学院に所属する研究者の研究テーマは多種多様である。結果としてデータの種類の種類は多種多様であり、ポリシーレベルで具体的かつ個別に指定することは、煩雑かつ、各部署単位での運用を限定する可能性がある。従ってデータの具体的な種類については言及せず、対象となるデータの範囲のみ提示することとする。データの由来に着目し、下記の通りに公開対象となるデータの分類をした。

・(A) 公開対象となるデータの定義、範囲

- ・ 機構の施設・設備を利用して得られたデータ
- ・ 外部の公的資金を活用して実施した事業・研究を通して得られたデータ

¹ 第5期科学技術基本計画 第4章 科学技術イノベーションの基盤的な力の強化

(2) 知の基盤の強化 ③ オープンサイエンスの推進(抜粋)

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。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

- ・ 外部組織との協業、共同研究等を通して得られ、公開に同意されたデータ
- ・ 及び、上記から得られたデータをもとに派生して得られたデータ

・(B) 潜在的に対象となるが一般的に公開せず、審査・契約等締結後に限定公開するもの

- ・ 機微な個人情報を含むデータ
- ・ 商業利用等に制約を課せられたデータ
- ・ 公的機関、国内の研究者等に、開示対象を制限すべきデータ

・(C) 対象から外すデータの定義・範囲

- ・ 国家安全保障、個人の安全・プライバシーに係るデータ

但し、個人の安全、プライバシーに関しては、適切な匿名加工処理を加えて公開することを検討し、それでも技術的、制度的に困難なものであると判断したもののみに制限する。技術的困難とは、安全性を担保した匿名加工後にデータとしての有用性を確保できないものである。制度的困難とは、そもそもその存在を秘匿すべきもの、外部への提供が許可されていないものである。

C.2.3 対象となるメタデータ

非公開となったものであっても、保有していることの情報開示をするために、メタデータは上記の公開の対象、非対象に関わらず、全てメタデータを作成し公開するものとする。但し、国家安全保障に係るもの

基本姿勢とする。その他の研究成果としての研究二次データについても、分野により研究データの保存と共有方法が異なることを念頭に置いた上で可能な範囲で公開する。ただし、研究成果のうち、国家安全保障等に係るデータ、商業目的で収集されたデータなどは公開適用対象外とする。また、データへのアクセスやデータの利用には、個人のプライバシー保護、財産的価値のある成果物の保護の観点から制限事項を設ける

及び院長から具体的に指定のあったものについては、メタデータも非公開の対象となる。非公開対象データについては、行政開示請求の文脈で対応することとする。

メタデータとは、検索システムの対象となるデータに関する情報である。しかし、検索システムによってメタデータの構造が異なるとメタデータの交換や横断的な検索が困難となるため、メタデータの構造を標準化する動きがあり、例えば書誌情報では書誌名、著者、日付、出版社等の標準的な15要素を定めた Dublin Core[3]が知られている。図書以外のデータについてのメタデータについても検討がなされており、採用するメタデータについては後述する。

以下、データ、メタデータの対象について整理した表を掲示する。

メタデータは基本的に公開するが、存在自体を知られるべきではないデータについてはメタデータも非公開とする。メタデータはリポジトリや検索システムに登録され、利用者が検索した時に検索結果として表示される。しかし、データの本体がリポジトリに登録されていないければ、データ本体へのアクセスは不可能である。このような状態が発生するのが、「公開範囲」が限定的公開・非公開に設定されている状況である。個人情報が含まれており、匿名加工がなされるべきものについては、匿名加工の処理が終了して公開の許可が下りたときにメタデータとデータ本体が同時に公開される。

匿名加工が困難あるいは加工するとデータの有用性が著しく落ちる場合は、限定的公開（メタデータのみ公開）での運用を検討する。

エンバーゴ期間（後述）が指定されているものについてはメタデータを先に公開し、データ本体はエンバーゴ期間経過後にアップロードして公開するか、リポジトリの機能を利用してエンバーゴ期間経過後に自動公開する設定を適用する。

なお、機微な情報を含むデータの公開に関する判断は ANDS のガイドラインで解説されており、メタデータとデータを公開するパターンの組合せを検討する際に参考にした[4]。

エンバーゴ (Embargo) については、複数の意味で使われており、どの文脈で使用されているのかを区別する必要があることに留意されたい。

(1) 論文公開前の広報（情報解禁日）としての意味

出版社側から論文公開前に論文に関するプレスリリースを著作者・機関側で出すことを自粛するよう要請しているもの。例えば、Nature research press の Embargo に関するポリシーは下記の通りである。

Nature research Press and embargo policies

<https://www.nature.com/nature-research/editorial-policies/press-and-embargo-policies>

表 公開対象とメタデータ、データの関係

We strongly discourage authors and

公開範囲	一般公開 (個人情報のないもの)	一般公開 (個人情報があるが匿名加工がなされたもの)	限定的公開	非公開	非公開 (機密)
メタデータ	公開	公開	公開	公開	非公開
データ	公開 (エンバーゴ指定されているものは期限後に公開)	匿名化後公開	非公開・契約締結後に公開	非公開	非公開

potential authors from direct solicitation of media coverage of material they have submitted to Nature and the Nature Research journals. Accepted contributions can be discussed with the media only once the publication date has been confirmed and no more than a week before the publication date under our embargo conditions. Please refer to the "Communications between scientists" section for more information about our embargo policy as it pertains to conference presentations and preprints.

このように期間を指定されたものについては、その期間以前にメタデータやデータをリポジトリに置いて公開、露出状態にしないことが求められる。

(2) 公開猶予期間

無料公開されない購読型論文が、一定期間を経て無料公開される際に設定される期間。ただし、無料公開されたからといって、その論文をリポジトリにても公開してよいとは限らないので、出版社の規約を確認されたい。

(3) 著者の権利や契約にもとづくもの

例えば、データを公開する研究者が、研究の先行利益を確保するために、リポジトリに登録してもただちに公開を希望しない場合に、公開予定日を設定する。リポジトリの機能によって公開予定日を過ぎた時に初めてデータが公開されるといった設定が可能である。

C.3 研究データの保存・管理・運用・セキュリティについて

・ 研究データの特性に応じたデータの保管、運用方針と国研としての取組について記述する。

(記述上の留意点)

・ 機関内で実施される研究活動におい

て順守すべき研究データの保存・管理・運用・セキュリティに関する対応についての方針、及びこれらを実施するための体制、並びにワークフローについて記述する。その際、研究データの特性、運用のフォローアップ、その他のポリシーとの整合性に留意する。

・ 研究データを登録するリポジトリ等について記述する。なお、特定のリポジトリ等名のほか、リポジトリ等が備えるべき条件について記述することが望ましい。

・ 研究プロジェクト終了後における研究データの保存・管理等の継続性にも考慮することが望ましい。

C.3.1 運用管理規程について

機関内で実施される研究活動において遵守すべき研究データの保存・管理・運用・セキュリティについては、既存の機関のセキュリティポリシー、システム運用管理規程に従うこととし、研究者の業務の増加を最小限に抑える。但し、行政文書についての分類やライフサイクルについての具体的な規程はあるが、研究、教育に関するデータについては、研究者の自主的管理に依存するところが多く、具体的な規程に乏しい。状況を調査し、必要に応じて規程を整備する必要がある。別途記述するようにデータ公開にそなえて、データを保管する期間・場所を定めることになるが、このデータ保管・管理を研究者に要求するのは業務圧迫につながる怖れがある。また人事異動、退職、研究プロジェクト終了等に伴い、データ管理の継続性が損なわれる可能性があることから、個人従属的な業務形態にすることは避けるべきである。従って、機関内あるいは「機関横断的」にデータのアーカイブに関わる担当者の設置を検討し、研究者からデータを引き渡されたあとは、管理責任は機関が一義に負うような体制、管理規程を整えることが望ましいと考えられる。

「機関横断的」とは、省庁下の国研群にお

いて、各々でデータを管理する担当者を雇用するのではなく、省庁単位で国研群のデータを一括して管理する部署・担当者を設置するアプローチである。現在、我が国においてはデータを管理できるスキルを備えた人材が不足しており、また雇用が不安定な傾向がある。長期的視野に立てば、専門スキルを持つ方を集約して安定して雇用できるような環境を創出することが望ましいと思われる。

C.3.2リポジトリ等が備えるべき要件等について

オープンアクセスにむけた研究データの管理方式は、Open Archive (OA) ジャーナルに投稿する Gold OA 方式と、機関リポジトリ等にセルフ・アーカイビングする Green OA 方式がある[5]。投稿した論文誌の出版社のポリシー、研究助成機関による論文の公開に関する方針もあり、出版された論文をセルフ・アーカイブすることが必ずしも認められていないため、どちらかの方式のみ採用するという事は不可能である。ただ、Gold OA の掲載にはコストがかかり研究者の負担、引いては税金の有効活用の観点から、可能な限り Green OA への掲載も認める出版社への投稿を推奨し、同時に機関におけるセルフアーカイビングを推進することが望ましいと考える[6]。また、今後の研究資金提供者によって OA セルフアーカイビングが義務つけられるようになった時に研究者を支援すべくセルフアーカイブが無償でできる機関リポジトリの存在は必要になると思われる。

Green OA のリポジトリの種類は、大学・研究機関みずからが保有する機関リポジトリ、研究者コミュニティによって運用される分野別のリポジトリ (arXiv 等)、助成機関によって運営される、助成を受けた研究成果を格納するセントラルリポジトリ (PubMed Central (PMC) 等) の形態がある[7]。本データポリシーの検討の対象となっているのは国立研究開発法人であり、国の事業

として実施している研究、調査にもとづいて開示すべきデータ等もあることが予想されることから、研究者コミュニティや助成機関等の第三者が運用しているリポジトリで開示するのではなく、自らあるいは委託したりリポジトリで開示するという運用を採るのが適切なケースがあると思われる。なお、NIH が Funding Agency として助成した研究の成果を公開させるために PMC をセントラルレポジトリとして提供しているように、国立研究開発法人が Funding Agency としての役割を果たしているならば、セントラルレポジトリの運用も担うことを検討すべきであると考えられる。

ただし、現状の研究機関の予算と人員状況下では、Green OA を実現するための体制を構築することは困難である。他の国立機関によって提供されているリポジトリ等のサービスを活用し、研究機関内にそのサービスを管理したり研究者によるデータ登録を支援したりする人員を配置、育成していくという方針を採用すれば、持続性のある運用が可能であると考えられる。もし、研究機関ごとに担当者を配置することが困難であれば、厚生労働省あるいは厚生労働省から指定された国立研究開発法人に人材を集約し、他の研究開発法人のデータ管理をもまとめて引き受けるとことで運用の最適化を図る必要があるかもしれない。

機関リポジトリのサービスについては、自前で機関リポジトリを構築する他、J-Stage や CiNii 等の外部リポジトリが多数使われている[8]。ただ、自前で機関リポジトリといっても完全に独自に構築している事例は少なく、外部の機関リポジトリサービスを利用しているのが実情である。国立研究開発法人の機関リポジトリとしては、オープンアクセスリポジトリ推進協会 (JPCOAR) [9] 下に、国立情報学研究所が運営している Japanese Institutional Repositories Online Cloud (JAIRO Cloud) という、機関リポジトリ環境を提供するサービスが検討に値すると考える。理由は下

記の通りである。

(a) 低廉なシステム運用コスト - 事業継続性の確保

JAIRO Cloud の利用は、JPCOAR 基本会費と JAIRO Cloud 利用料金の組合せの年間会費を支払えば可能であり、構成員が 200 名規模では年間 10 万円程度、1000 名規模でも数十万円程度と低廉な利用料金の設定がなされている(2019 年度時点)。また、JPCOAR 会員に加入するため、JPCOAR で開催される人材育成や啓蒙活動に参加することで、独自に取り組むよりもオープンデータに関わる人材育成のマネジメントの負担も軽減されることが期待される。

(b) 科学技術基本計画のプロットへの追従

JPCOAR は、2019 年～2021 年度にかけての戦略[10]として、JPCOAR はオープンサイエンスの推進に寄与するため、研究データの公開、流通に関する先導的な取組み、オープンアクセスを推進する学術情報流通の基盤を整備し、コンテンツの流通、活用を促進する、オープンアクセス、オープンサイエンスの推進に対応できる人材育成を行う、ことを挙げている。JPCOAR の会員となり、活動に参加することで最新のオープンサイエンスに関する動向の把握と、最新の機関リポジトリ環境の整備、人材育成に寄与することが期待される。

(c) 外部への委託による機関内業務の削減

国立情報学研究所によって情報セキュリティ、システムの継続的運用上の配慮が十分になされたクラウド環境に機関リポジトリ業務を委託することにより、機関におけるシステム管理者の雇用、情報セキュリティの確保に係る諸業務を節約できる。また、クラウドにデータを展開することにより、自然災害、停電等によるデータの損壊、損失の機会を最小化し、オープンサイエンス時代に求められているデータ保存・管理の継続性に対する要件に対応することが可

能になる。すなわち、DPG で求められている「研究プロジェクト終了後における研究データの保存・管理等の継続性への考慮」にも応えられると考える。

(d) データの真正性、保存性の確保

データの保存期間は事業が継続する限り、半永久的としたい。保存に関するコストへの配慮もそうであるが、長期間保存によるデータの消失、損壊を回避する必要がある。オンプレミスの運用であればサーバやストレージの更新によるマイグレーションが必要であるが、データ移行はデータの欠損、破壊の最も大きな契機の一つであり可能な限りさけたい。また、ストレージに高品位なものを利用することでデータそのものが損傷を受ける機会を削減したい。このような要求を鑑みれば、データの多重複製、データ損傷のモニタリング、自動修正を行う高機能なストレージに保管することが望ましい。近年のクラウドサービスはビッグデータを確実に保存するためにオブジェクトストレージ等の仕組みを取り入れてデータ保存に関する品質を飛躍的に高めている。そういう意味でも最先端のストレージ技術の運用が期待されるクラウドベースでの運用としたい。

なお、DPG には「実施するための体制、並びにワークフローについて記述する。」という言及はあるが、これは院内で運用される対策基準あるいは実施手順レベルにおいて記述すべきものであり、対外的に公開されるポリシーにおいて記載するような内容(粒度)ではないと考える。

C.3.3 研究データのマネジメントについて

研究データの管理は研究を開始する前に計画すべきであり、標準的な研究上の実務に組み込まれるのであれば、必ずしも時間や費用を大きく費やすことにはならないという主張がある[11]。しかし、現時点

では我が国の Funding Agency でも Research Data Management (RDM) の提出を求めているが、必ずしも実際の研究上の詳細な計画まで踏みこんだ内容になってお

エビデンスについて確認されていない。研究データ管理用基盤の導入について検討するのはよいとしても、この数年間以内に研究データ管理用基盤を利用したマネジメン

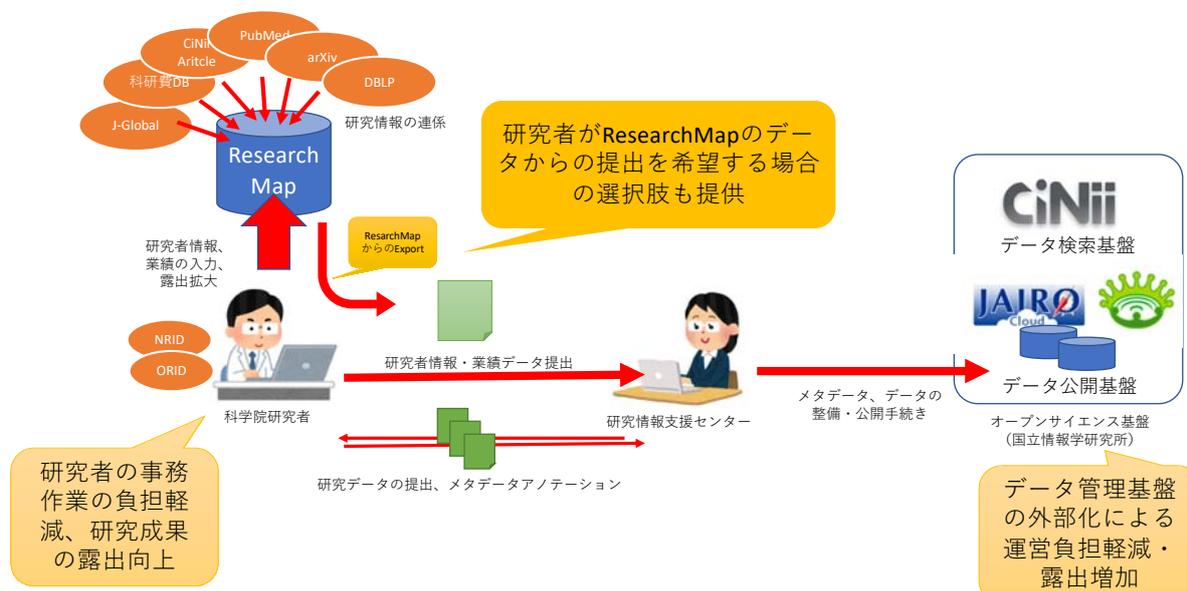


図1 研究データベースとリポジトリの連携を活用したオープンデータ登録支援

らず、また研究者が利用できる標準的なデータ管理環境も存在しておらず、依然として研究者自身による管理に委ねられている。

我が国では、国立情報学研究所オープンサイエンス基盤研究センターが米国 NPO の Center for Open Science が開発している研究データ管理用の Open Science Framework[12] をカスタマイズして GakuNin RDM[13] という研究データ管理基盤を提供している。研究データ管理基盤をあらためて導入するメリットは、研究者の本来の業務ではない研究データを安全に管理するためのシステムの運用・保守を外部化できること、データの永続性を保証し研究資料の消失・散逸を抑制するというセキュリティ保全[14]の他に、研究者コミュニティでのデータ共有を促進し、研究プロセスを共有化・再利用することで研究を加速する効果が期待される。

ただ、この GakuNin RDM を用いた先行検証を含めて、一般的に利用出来るかたちで提供されている研究データ管理用基盤を用いた研究活動およびその成果として得られた

トが確立されている状態を想定するのは困難である。

従って、当面は RDM システムの活用に向けた取り組みと並行して、従来のプロセスで出来上がったデータをどのように公開するかを検討を進めることとしたい。

対象となる研究論文、データが発生したあと、どのようなプロセスを経て公開に至るかを検討する必要がある。現実的な問題として、対象論文・データの把握は研究者の自己申請に依らざるを得ない。自己申請に関して、研究者やデータ公開を支援する担当者の業務負担を軽減するために下記の様な運用モデルの検討を進めているところである。

別途に記述するように、研究者には国際的な研究者 ID の獲得と研究者データベースへの業績登録を依頼する。その代わりに、これまで年度の事業報告や業績評価の為に作成していた資料から研究、論文に関する業績の報告にかかる業務を免除するようなサポートをする。具体的には、研究機関の

データ公開支援者は、研究機関に所属している研究者が登録している研究者データベースをクローリングし、新たな業績が登録されたことを確認次第、データ公開申請書を起票し、当該業績に関する書誌情報等を予め埋めた上で、研究者に公開の可否、成果物の提出について申請させる Web サイトへの案内をする。例えば、研究者データベースについては、国立研究開発法人科学技術振興機構が提供している次世代研究基盤リサーチマップ(researchmap) [15]によって、各種データベースと連携し、各種研究補助金の書類や評価データの提出フォーマットをダウンロード可能なサービスが提供されている。また、researchmap のデータベースからデータを入手する Web サービスも提供されている。この Web サービスを利用して複数の研究者に関する情報を XML 形式で一括して入手可能とされている。この XML データから研究機関が使用しているリポジトリサービスに登録するためのメタデータの雛形を自動生成する仕組みを開発することで、データ公開に関する業務効率化が期待される。論文データベースは DBLP、PubMed、ORCID、Web of Science 等多数あるが、researchmap に注目する理由は、複数の論文データベースや書誌情報と連携して業績を取り込めること、科研費や業績報告など、我が国に関する研究者の各業務を支援する機能が提供されている等、研究者にとっても機関にとっても業績の管理をワンストップで提供するプラットフォームであると考えている。

C.3.4 TODO

- ・来年度以降、データポリシーにもとづいた運用を開始する以前に、研究データの管理にかかわる情報システムの管理規程等の確認・整備が必要である。

- ・Jairo Cloud 等、外部のリポジトリサービスを利用してリポジトリを運用する場合は、保有データの外部組織への管理委託に関する規程の確認が必要である。

- ・researchmap の研究データベースから入

手できる情報とリポジトリに登録するメタデータについての調査とコンバートプログラムの開発

C.4 研究データに対するメタデータ、識別子の付与、フォーマットについて

- ・研究データに対するメタデータ及び識別子付与についての方針を記述する。また、研究データの特性に応じた標準的なフォーマットが存在する場合は、それも併せて記述する。

C.4.1 取り扱うメタデータの規格

国際的なデータ公開、再利用を推進するために、研究データに対するメタデータ、識別子は国際的に普及している規格を優先的に採用するものとする。

我が国では、機関リポジトリの標準的なメタデータスキーマとして junii2[16]が流通しており、リポジトリ登録に使用されている。しかしながら、研究データの記述や OA 状況をモニタリングするために必要な要素、各種識別子を記述するための要素を十分に備えていないとされる[17]。国際的には研究データのメタデータとして DataCite のメタデータスキーマ[18]が標準的なものとして定着しており、junii2 も DataCite のメタデータとのハーモナイズが検討されていた[17]。この流れをうけて、junii2 に代わるメタデータスキーマとして 2017 年 10 月に JPCOAR スキーマ ver 1.0、そして 2018 年 8 月に現時点で最新版となる ver1.0.1 がリリースされている[19]。Confederation of Open access Repositories (COAR)や DataCite で利用されている統制語彙を当面必要とする用途に絞って採用し、研究データ等の新しいコンテンツへの対応、OA の状態 (OA かどうか、エンバゴ終了日など)、公的研究助成に関する管理情報、多様な論文の種類を網羅するなど、JPCOAR コアスキーマは学術情報の国

際的な流通性を高め、かつオープンサイエンスに対応可能なメタデータの設計がなされている[20]。我々は国際的な研究機関とも協調して事業にあたることを期待されているから、このような国際的な流通性に配慮したメタデータを積極的に採用すべきであると考え。また、共用リポジトリサービスの JAIRO Cloud では2019年度に JPCOAR スキーマに対応した JAIRO Cloud を運用開始予定であるとのことである[20]。

TODO:[17]によれば、現在の機関リポジトリは研究データの DOI 付与には対応していないとのこと。現在の機関リポジトリの実装について状況確認。NII 等の共同研究を検討し、機関リポジトリに必要な要件の確認と機能の開発に協力する。

TODO:JPCOAR メタデータ規格について検討し、研究者の保有データ調査の為の基礎資料とする。(分類手法など)

C.4.2 研究者の識別子

著者を管理する ID スキームは複数提案されている。研究者の負担軽減のために、シェアの大きい ID スキームの採用を検討する。日本における研究助成申請時に使われる科研費研究者番号をベースとした NRID と、国際的な研究者識別子付与活動をリードしている ORCID [21] (Open Researcher and Contributor ID) を中軸として利用することが望ましいのではないかと。また、科学研究助成事業データベース上で NRID から ORCID のリンケージが可能となっている[22]。研究機関全体の取り組みとして、着任時の研究者向け研修(図書館、論文データベースの操作研修等)等の機会を利用して、ORCID の取得、そして NRID と ORCID のリンケージ設定を推奨するのがよいのではないかと。

メモ:

・NRID: 科研研究者番号 KAKEN データベース等で使用

・e-Rad: 府省共通研究開発システムで使われる研究者用 ID

・ORCID 世界中で研究者を一意に特定するための研究者用 ID 非営利団体 ORCID, Inc によって運用されている。

C.4.3 組織の識別子

科研費電子申請システムで利用されている機関番号[23]と世界の研究機関データベース GRID (Global Research Identifier Database) [24]を採用するものとする。また、2019年2月に CrossRef から GRID をベースとした Research Online Registry (RoR) が発表されており、こちらの動向も注目していきたい。

例) 国立保健医療科学院の GRID は grid.415776.6

RoR は <https://ror.org/0024aa414>

なお、識別子については採用するメタデータ規格の入力ガイドライン等で推奨されているものが存在する可能性があるため、メタデータ規格の検討時に識別子についてもあらためて検討することとする。

C.4.4 データフォーマットについて

標準形式への研究データの準拠については、データを生成した部局において個別に判断するものとする。既存フォーマットを標準規格に準拠した別のフォーマットに変換することは追加のコストや労力がかかる可能性があり、利用者側のニーズ(何のデータをどのような形式で利用したいか等)が把握出来ていない現状では優先度は低く、まずはデータの存在について周知していく取り組みについて優先度を割り振るべきであると思われる。

但し、再利用性が高い状態で公開することを要請されているため、ポリシーでは原則としてスター・スキーム[25]の3段階目

以上の標準的形式での公開を目指すことにしたい。

例えば、国立保健医療科学院は公衆衛生分野を主に担当するが、公衆衛生分野では、近年の国境を越えた人や資源の移動にともなう自然環境、感染症などについて対策がボーダレス化しつつある今、データの迅速な公開、共有にむけて標準規格の準拠についての重要性が認識されつつある。しかし、分野によってはデータの構造についてコンセンサスが得られている標準規格は殆ど存在しない。各分野においてのデータの標準化に関するベストプラクティスが蓄積され、相互運用性の確保の努力が行われることを期待し、現時点では「当研究分野において標準化された規格があれば、可及的にそれに準拠したデータ形式で公開すること」と述べるのに留めるのが望ましいと考える。

なお、Resource Definition Framework (RDF) [26]及び、RDF によって記述された Semantic Network からデータを抽出する SPARQL クエリ言語がオープンかつ標準的なデータとして公開する手法として知られているが、結局のところ当該研究ドメインに関するリソースの表現を統一するための語彙定義がなされていることが前提であり、先述したように公衆衛生分野では発展途上である。所謂「標準規格に準拠したデータ」のありかたについては、理想と実利のバランスをとりつつ、下記の取り組みを並行して進めるのが望ましいと思われる。なお、前述した「スター・スキームの3段階目以上の標準的形式」に関しては、以下の条件を基準に選定することが望ましいと思われる。

- ・長期的な互換性を確保することを努力しているソフトウェアで使われている形式であること。

全てのデータはデータアクセスに関するハードウェアおよびソフトウェア環境が

旧式化するというリスクにさらされている [11]。そのため、過去のバージョンで作成されたデータをインポートする下位互換性を確保し、長期的なデータアクセスを保証することをサポートしているソフトウェアがあるならば、そのソフトウェアが利用している形式を優先して採用することを検討する。

例えば PDF 形式文書はスター・スキームで推奨されているフォーマットではないが、やむをえず利用する場合は 長期保存のための国際規格 ISO 19005-1(PDF/A)に準拠した PDF 形式で保存する等の配慮が必要である。これは長期にわたって保存しても表示される内容、色等が変化しないで再生できる PDF 文書を提供することを目標として策定された規格である。

- ・データ形式に関する仕様が公開されていること。

ソフトウェアがオープンソースのものであればソフトウェアの開発が停止してもソースコード等からデータ形式の仕様を確認し他のデータ形式に移植する可能性が残されている。しかし、商用ソフトウェアではサポート中止後、仕様が開示されずブラックボックス化し、データへのアクセスが困難になることが考えられる。現時点でデータ形式に関する仕様が容易に入手できるものを選択することが望ましい。但し、これは商用ソフトウェアの仕様を排除するものではない。例えば、統計処理ソフトウェア SPSS では公式の公開仕様は存在しないが、GNU PSPP プロジェクトより SPSS Portable Format(.por)²、SPSS System file(.sav)³の仕様が公開されており、オープンソースの統計処理ソフトウェア R[27]の foreign パッケージや機械学習等に多用されている開発言語 python の pyreadstat ライブラリでサポートされている。同様に統計処理ソフト SAS、Stata、地理情報空間システムに使われる地

² https://www.gnu.org/software/pspp/pspp-dev/html_node/Portable-File-Format.html#Portable-File-Format

³ https://www.gnu.org/software/pspp/pspp-dev/html_node/System-File-Format.html

理空間データ ESRI シェイプファイル等、プロプライエタリなフォーマットであるが仕様が公開されており、多数の汎用的なツールでサポートしているものは受け容れても良いと思われる。

Excel や OpenOffice のスプレッドシートソフトウェアも Open Document Format for Office Applications (ODF) フォーマットで汎用的なかたちでデータを記述できるようになっているが、Excel は表現力・自由度が高く表や罫線、セルの加工が容易に出来るため、結果としてデータの二次利用に適さないかたちでの公開につながる可能性が大きく [28]、データの二次利用性向上にむけたガイドラインを定義することも困難なことから極力使用しないことが望ましいと思われる。

(i) 当座のデータ公開にむけて

最終的には Semantic Web のフレームワークの文脈に従ってデータを公開することが望ましいとされている。そのため、現時点では Semantic Web に則って公開できなくても、現在どのレベルにあるのかを自覚的に認識し、進歩の方向性を見定めることは有用である。オープンデータとして公開するときに、ティム・バーナーズ・リーはスター・スキーム (star scheme) として 5 段階の構造化レベルを考慮し、最終段階である Link of Data (LOD) を意識していくことを提案した [29]。

- ★ Available on the web (whatever format) but with an open license, to be Open Data
- ★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)
- ★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that

people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

図 Five Star Scheme Linked Data - Is your Linked Open Data 5 Star?より引用⁴

それによれば、一つ星はどのような形態であれ、オープンライセンスで公開されていること、二つ星は機械可読な構造化データとして利用可能 (Excel 等が例示されている) にすること、三つ星は二つ星に加えて非独自フォーマット、すなわち特定のベンダー製品の使用を要求しない (CSV 等) もにすること、四つ星はさらに W3C の標準規格である RDF と SPARQL を使用して記述すること、五つ星は、これらのデータが他のデータとリンクされる状態で公開されている状態である。オープンサイエンスに資する観点からは、データの再利用の障壁を下げると同時に研究者に負担がかからないような運用が望ましく、最低三つ星の水準を満たすかたちでの公開を求められれば十分であろう。公開データ形式選択にあたっての基準と事例については前項で考察している。

ただ、スタースキームでは機械可読可能なフォーマットに加えて非独自フォーマット、すなわちベンダー製品の使用を要求しないものを、より上質なデータとして位置づけている。このような論理では商用統計処理ソフトよりベンダー非依存の CSV 形式の方がよいという判断になる。しかしながら、3つの理由から必ずしもそのような選択が最善とは限らないと思われる。一つ目は先述したとおりプロプライエタリーのものであっても仕様が公開され多くのツールから利用可能な状態であり、長期に渡って情報資産として活用できるものであれば容認しうると考える。もう一つの理由はより積極的なものであり、データそのもの以外にデータ構造、欠損値定義、ラベル、コメント等のメタデータがある場合、CSV に変換す

⁴ <https://www.w3.org/DesignIssues/LinkedData.html>

ることによってそれらの一部が失われる可能性があることである。そして最後の理由は、CSV はデータ構造が固定的であり、構造の変化に弱く長期に渡るデータの互換性を確保することが困難であることから、長期間にわたるプロジェクトやデータ構造の要件変更が発生するようなものには向いていない可能性がある。最終的なセマンティックスペースでのデータ処理に対応できるよう、大元のデータおよびメタデータの形態が最大限に保存され、なおかつ長期的にアクセスが保証されている形式を選択することが望ましいと思われる。

(ii) データ構造の標準化にむけて

公衆衛生分野でのデータの標準化は2つの方向性が考えられる。

一つは規制当局や国の事業のための情報収集を用途とし、二次利用性を高めるために内容の標準化をはかることである。CDISC がそのような用途の為の標準化をすすめており、この CDISC の SDTM において公衆衛生分野の語彙を開発、導入することに寄与していくことが考えられる[30]。

もう一つは、LOD/RDF の文脈にあわせて統計データとして公開するためのしくみとして、SDMX (Statistical Data and Metadata eXchange) [31] をベースとした RDF データキューブ [32] の形態での公開を目指すことである [33]。

いずれにしても、公衆衛生分野は LOD の事例は立ち後れている分野なので、研究データを保有している研究者と協業し、公衆衛生分野の LOD を構築する研究プロジェクトと予算獲得にむけた画策は必要であろう。

C.5 研究データの帰属、知的財産の取り扱いについて

- 研究データの帰属及び知的財産の取り扱いについて、国研の関係規程を踏まえた上で、研究データの利活用の方針に応じて記述する。この記述は、保管に際して遵守すべきルールとして規定するとともに、同ルールと研究データ利活用のルールと整合を取る。
- 研究データに係る作成者、管理者等の免責事項について記述する。

公的な研究資金にもとづいた研究活動、あるいは機構の施設・設備等を利用した研究、業務の過程で取得されたデータは、当該研究者との特別な取り決めがない限り、機構に帰属するものとされている。

C.5.1 データを公開する際のライセンスについて

データを公開するにあたり、基本的には FAIR の原則にもっとも適うライセンス形態として推奨されているクリエイティブコモンズの CC BY 4.0 に準拠して公開することが望ましいと考えられる。これは、オリジナルから改変された派生物が共有されることと、商用利用を許容するものである。

日本政府は各省庁が保有しているデータをオープンデータとして公開すべく、DATA GO. JP というデータカタログサイト⁵を提供しており、その利用規約では CC BY 4.0 国際でのライセンスが指定されている。

いわゆる「オープン」なライセンスにはオープンソースのライセンス群が知られているが、こちらは主にプログラム、ソースコードに適用されるものであり（もちろん、研究の過程で作成されたアルゴリズム、プログラムを公表するのであれば、オープンソースでの適用となろう）、データや論文にはそぐわないものである。

⁵ <https://www.data.go.jp/>

データに関して「オープン」なライセンスには同じくクリエイティブコモンズのCC0⁶、オープンデータ・コモンズのODC-BY⁷が知られている。但し、CC0は所謂著作権を放棄するPublic Domainであり、我が国では著作権の放棄に関する明文の規定は存在せず、国立研究開発法人が適用するライセンスとしては不適切であると思われる。

ただ、メタデータに関する検討でも議論したように、機微な情報を含むデータを扱うこともあり、ただちにオープンできるデータのみではなく、個別に検討する必要がある。そのため、データポリシーにおいては一律のライセンスを指定するのではなく、FAIRの原則に則したライセンスを優先的に採用することを記述するにとどめることが望ましいと考える。

C.5.2 商業利用への提供について

独立行政法人海洋研究開発機構のデータポリシー[34]では、産業利用については原則として有償として、適切な対価を徴収することが記載されている。本稿での検討の対象となっている国立研究開発法人においては、民間より何らかの対価を徴収するような運用についての検討に乏しいところがある。文部科学省の「研究開発成果としての有体物の取扱いに関するガイドライン⁸」において、成果有体物については商業利用を目的とした者への提供の場合は、「提供を要請する者と各機関との間において、成果有体物の取扱いに関する必要な条件を明記した譲渡又は貸付契約を締結し、有償で提供する。」とあるとおり、有償での提供が一般的な扱いとなっている。一方、厚生労働省においてはそのようなガイドラインは現存せず、なおかつ法人化された国公立大学とも異なり、国の試験研究機関は現在でも

6 <https://creativecommons.org/publicdomain/zero/1.0/>

7 <https://opendatacommons.org/licenses/by/summary/>

8 http://www.mext.go.jp/a_menu/shinkou/sangaku/sangakuc/020901.htm

「国」扱いのため、個別の検討が必要である。データポリシー策定と並行して、職務発明規程や就業規則の修正・追補を検討したい。

ただ、商用利用に関する制約を入れることは、CCにおける「オープン」の定義[35]に反する可能性がある⁹ため、先に推奨ライセンスとして提案したCC BY 4.0と衝突しうることも念頭におかなければならない。

C.6 研究データの公開、非公開及び猶予期間並びに引用について

・研究データの公開について、機関の研究データの利活用の方針に応じてデータ公開までの猶予期間を適切に設定し、それに基づく公開時期について記述する。

・公開データの利用に際しては、利用者に対して適切な引用を求める。その際、識別子を用いた引用情報の記載ルールを設けるなど、他のユーザーが引用元のデータを参照できるよう配慮する。

研究データは出版社や研究者自身によって設定されたエンバーゴ期間を過ぎれば、一般公開に支障がない限り遅滞なく公開するものとする。一部の出版社や学協会では出版後一定期間が経過するまで、Green OAで同一論文を公開することを認めない(embargo: 公開猶予期間)が存在することがある。また研究者によって合理的かつ公益を損ねない範囲での研究の競争上の優位性を保つための公開までの猶予期間の指定がある場合は、その猶予期間の指定後にデータを公開するものとする。研究者のモチベーション維持のために、公益を損ねない範囲での研究者の利益を擁護することは必要である。

9 2.1.9 料金領収の禁止 ライセンスは、その条件の一部としていかなる料金支払いの取り決めや、ロイヤリティ、あるいはその他の補償行為あるいは金銭的代償を規定してはならない。

公開データの利用に際しては、利用者に対して適切な引用を求めることとする。但し、リポジトリによるデータ識別子の永続的な管理体制とデータ引用の標準様式が普及していないため、当面は学術論文の投稿規程等で定められた様式で引用することを要求するところから始めるものとする。リポジトリを構築し、データに対して安定した識別子が付与されるようになれば、識別子を用いた引用情報の記載ルールを設けることとする。

D. 参考文献

1. 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. *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*. 2018; Available from: <https://www8.cao.go.jp/cstp/stsonota/datapolicy/datapolicy.pdf>.
2. 内閣府 政策統括官（科学技術・イノベーション担当）. *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～解説資料～*. 2019; Available from: <https://www8.cao.go.jp/cstp/stsonota/datapolicy/dpguideline.pdf>.
3. Weibel, S., et al., *Dublin core metadata for resource discovery*. Internet Engineering Task Force RFC, 1998. 2413(222): p. 132.
4. Guides, A. *Publishing and sharing sensitive data*. 2018; Available from: http://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf.
5. Initiative, B. O. A., *Read the Budapest open access initiative*. Budapest Open Access Initiative, 2002.
6. Harnad, S., *Fast-forward on the green road to open access: the case against mixing up green and gold*. arXiv preprint cs/0503021, 2005.
7. 東京大学附属図書館, *オープンアクセスハンドブック 第2版*. 2017.
8. 課題領域：オープンサイエンス(SCPJ)班, 機. *国内学協会のオープンサイエンス対応状況調査(報告)*. 2016; Available from: <http://id.nii.ac.jp/1280/00000199/>.
9. オープンアクセスリポジトリ推進協会. 2019; Available from: <https://jpcoar.repo.nii.ac.jp/>.
10. JPCOAR, *JPCOAR オープンアクセスリポジトリ戦略 2019～2021年度*. 2019.
11. Van den Eynden, V., *Managing and sharing data: Best practice for researchers*. 2011: UK Data Archive.
12. Foster, E.D. and A. Deardorff, *Open science framework (OSF)*. Journal of the Medical Library Association: JMLA, 2017. 105(2): p. 203.
13. 込山悠介, *研究データ管理サービス: GakuNin RDM*. 2019.
14. 国立情報学研究所オープンサイエンス基盤研究センター. *GakuNin RDM (研究データ管理基盤)*. 2017; Available from: <https://rcos.nii.ac.jp/service/rdm/>.
15. 坪井, 彩. and 治. 大須賀, *JST サービス紹介 国内最大級の研究者総覧 researchmap*. 情報管理, 2018. 60(12): p. 906-909.
16. 学術機関リポジトリ構築連携支援事業. *メタデータ・フォーマット junii2 (バージョン 3.1)*. 2014; Available from: <https://www.nii.ac.jp/irp/archive/system/junii2.html>.
17. 大園, 隼., *サンメディアソリューションセミナー オープンサイエンスの最新情報: メタデータの相互運用性を中心に*. 薬学図書館 = Pharmaceutical library bulletin, 2017. 62(1): p. 40-47.

18. Group, D.M.W. *DataCite Metadata Schema 4.2*. 2019; Available from: https://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadataKernel_v4.2.pdf.
19. オープンアクセスリポジトリ推進協会. *JPCOAR スキーマガイドライン Ver 1.0.1*. 2018; Available from: <https://schema.irdb.nii.ac.jp/ja>.
20. 大園 隼彦, et al., *JPCOAR スキーマの策定：日本の学術成果の円滑な国際的流通を目指して*. 情報管理, 2018. 60(10): p. 719-729.
21. 蔵川, 圭. and 英. 武田, *研究者識別子 ORCID の取り組み*. 情報管理, 2012. 54(10): p. 622-631.
22. Kurakawa, K., et al., *Researcher Name Resolver: identifier management system for Japanese researchers*. International Journal on Digital Libraries, 2014. 14(1-2): p. 39-58.
23. 独立行政法人日本学術振興会. *機関番号一覧*. 2019; Available from: <https://www-shinsei.jsps.go.jp/kaken/topkakenhi/codelist-ka.html>.
24. Science, D. *GRID Global Research Identifier Database*. 2019; Available from: <https://www.grid.ac/>.
25. Berners-Lee, T., *Five star open data*. 2009.
26. Lassila, O. and R.R. Swick, *Resource description framework (RDF) model and syntax specification*. 1998.
27. R Core Team. *R: A Language and Environment for Statistical Computing*. 2019; Available from: <https://www.R-project.org/>.
28. 奥村 晴彦, 「ネ申 Excel」問題. 情報教育シンポジウム 2013 論文集, 2013. 2013(2): p. 93-98.
29. Tim Berners-Lee. *Linked Data*. 2010; Available from: <https://www.w3.org/DesignIssues/LinkedData.html>.
30. 木村 映善 and 上野 悟, *公衆衛生分野でのデータ利活用に貢献する標準医療情報規格と CDISC 標準*. 保健医療科学, 2019. 68(3): p. 212-218.
31. SDMX, S., *Statistical Data and Metadata Exchange*. URL: <http://sdmx.org>, 2011.
32. W3C. *The RDF Data Cube Vocabulary*. 2014; Available from: <https://www.w3.org/TR/vocab-data-cube/>.
33. 西村, 正., *Linked Open Data (LOD) による統計データの提供：政府統計データ (e-Stat) の新しい形*. 情報管理, 2017. 59(12): p. 812-821.
34. 独立行政法人海洋研究開発機構, *データ・サンプルの取り扱いに関する基本方針 (データポリシー)*. 2007.
35. Foundation, O.K. *Open Definition 2.1*. Available from: <http://opendefinition.org/od/2.1/ja/>.

研究成果の刊行に関する一覧表

該当なし

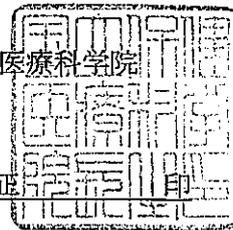
令和2年3月23日

厚生労働大臣 殿

機関名 国立保健医療科学院

所属研究機関長 職名 院長

氏名 福島 靖正



次の職員の令和元年度厚生労働行政推進調査事業費の調査研究における、倫理審査状況及び利益相反等の管理については以下のとおりです。

- 研究事業名 厚生労働科学特別研究事業
- 研究課題名 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
- 研究者名 (所属部局・職名) 統括研究官
(氏名・フリガナ) 木村 映善・キムラ エイゼン

4. 倫理審査の状況

	該当性の有無 有 無		左記で該当がある場合のみ記入 (※1)		
			審査済み	審査した機関	未審査 (※2)
ヒトゲノム・遺伝子解析研究に関する倫理指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
遺伝子治療等臨床研究に関する指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
人を対象とする医学系研究に関する倫理指針 (※3)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
その他、該当する倫理指針があれば記入すること (指針の名称:)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>

(※1) 当該研究者が当該研究を実施するに当たり遵守すべき倫理指針に関する倫理委員会の審査が済んでいる場合は、「審査済み」にチェックし一部若しくは全部の審査が完了していない場合は、「未審査」にチェックすること。

その他 (特記事項)

(※2) 未審査の場合は、その理由を記載すること。

(※3) 廃止前の「疫学研究に関する倫理指針」や「臨床研究に関する倫理指針」に準拠する場合は、当該項目に記入すること。

5. 厚生労働分野の研究活動における不正行為への対応について

研究倫理教育の受講状況	受講 <input checked="" type="checkbox"/> 未受講 <input type="checkbox"/>
-------------	---

6. 利益相反の管理

当研究機関におけるCOIの管理に関する規定の策定	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由:)
当研究機関におけるCOI委員会設置の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合は委託先機関:)
当研究に係るCOIについての報告・審査の有無	有 <input checked="" type="checkbox"/> 無 <input type="checkbox"/> (無の場合はその理由:)
当研究に係るCOIについての指導・管理の有無	有 <input type="checkbox"/> 無 <input checked="" type="checkbox"/> (有の場合はその内容:)

(留意事項) ・該当する□にチェックを入れること。
・分担研究者の所属する機関の長も作成すること。