

社会技術研究開発センター(RISTEX)の研究開発における研究データ管理の状況調査報告

令和6(2024)年8月

国立研究開発法人科学技術振興機構(JST)

社会技術研究開発センター(RISTEX)

1.はじめに

本稿は、国立研究開発法人科学技術振興機構 社会技術研究開発センター(以下、RISTEXとする)が、社会技術研究開発事業の関係者を対象に実施した、研究データ管理に関するオンラインアンケート調査の結果を報告するものである。

研究活動のデジタル・トランスフォーメーション(研究DX)の流れが世界的に加速し、オープンサイエンス、データ駆動型研究等の議論が活発化する現在、研究データ管理(Research Data Management, RDM)¹の重要性が増している。

研究者と現場のステークホルダーが連携して社会の問題の解決に資する研究開発を長年推進しているRISTEXのプロジェクトにおいては、個人情報やパーソナルデータを多く扱い、かつ、それらのデータを活用した社会実装を視野に入れていることから、より適切な研究データ管理が求められると考えられる。そこで、多様な分野の研究者と研究機関のみならず、研究に関与する様々な組織を巻き込んだデータ管理が必要であるとの問題意識を持ち、センターとして課題の検討を進めている²。

このような取り組みを進めるに際し、研究活動の現場に即した議論と対応を行うためには、社会技術研究開発事業における研究データ管理の状況を把握することが重要である。そこで、RISTEXの関係者が日常的にどのようなデータをどのように取り扱っているか、また、どのような要望を持っているかを把握することを目的として、アンケート調査を実施することとした。

本調査にあたり、次の調査を参照した。文部科学省 科学技術・学術政策研究所(NISTEP)が2016年から隔年で科学技術専門家ネットワークに所属する研究者に対して実施している「研究データ公開と研究データ管理に関する実態調査2022：日本におけるオープンサイエンスの現状³」(以下、NISTEP調査とする)、及び、オープンアクセスリポジトリ推進協会(JPCORE)と大学ICT推進協議会(AXIES)が2020年に大学や研究機関に対して調査を実施し、NISTEPが分析を行った「日本の研究機関における研究データ管理(RDM)の実践状況⁴」(以下、JPCORE/AXIES調査とする)である。これらの調査と本調査は調査対象や回答数等が異なるが、参照した調査項目については比較情報を参考として結果に記載する。なお、先行するこれらの調査は研究データ管理にかかわる広範な内容を調査しているが、本調査は上述のRISTEXの研究開発の特徴を踏まえ、特に研究開発プロジェクト内外のデータの入手と提供・共有にかかわる項目が多くなっている。

以下、まず調査の概要を示した後、結果を(1)取扱う研究データの種類・性質等、(2)研究データに関する所属機関との関係、(3)研究データの入手と提供・共有の3つのテーマに分けて示す。

2.調査の概要

調査は2024年5月7日～21日にウェブのアンケートシステムを用いて実施した。質問数は、必須項目の18問を含む、28問であった。RISTEXで活動中の5つの研究開発プログラム⁵を通じて各マネジメントチーム構成員（総括とアドバイザー）と研究開発プロジェクト参画者（研究代表者等）にeメールで調査案内を発出するとともに、活動が終了している研究開発プログラム及びプロジェクトのうちRISTEXのメーリングリスト登録者にも同様に案内を行い、有効回答68件を得た。

回答者の所属機関は、「国立大学」45.6%、「私立大学」27.9%と、大学が70%以上を占め、「企業」7.4%、「NPO・一般社団法人」7.4%、「国立研究開発法人」7.4%と続いた。年齢層は、「30代」8.8%、「40代」26.5%、「50代」25.0%、「60代」32.4%、「70代」7.4%であった。研究分野は13分野⁶を提示したところ、「社会科学」41.2%が最も多く、「医学」13.2%、「工学」13.2%、「心理学」8.8%、「計算機科学」7.4%と続いた。自身の活動については、「論文出版や学会発表などの学術研究活動を行っている」79.4%、「研究支援を行っている」8.8%、「実務者として研究協力を行っている」7.4%であった。さらに、これまで経験した研究のうちトランスディシプリナリー（TD）研究⁷の割合について質問したところ、「100%」5.9%、「80%～100%未満」8.8%、「60%～80%未満」14.7%、「40%～60%未満」16.2%、「20%～40%未満」20.6%、「1%～20%未満」25.0%、「0%」2.9%、「わからない・覚えていない」5.9%であった。

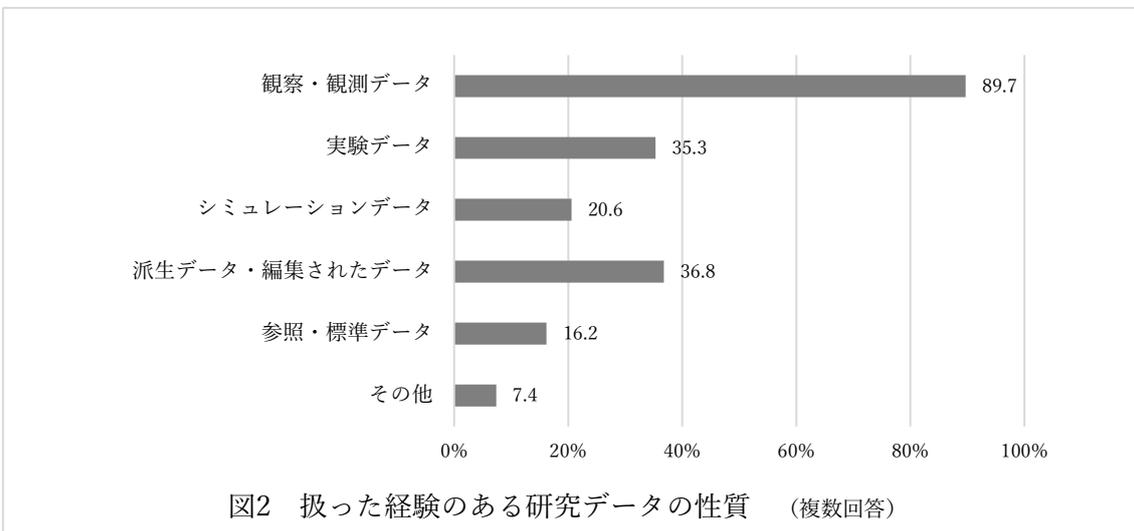
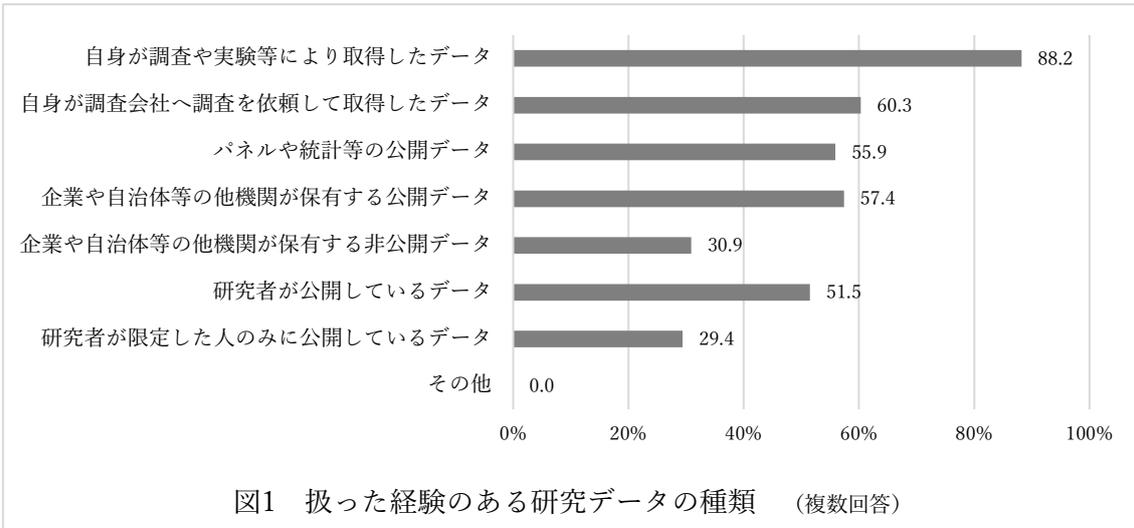
なお、回答者のRISTEXへの関与状況は、「研究開発プロジェクトに、研究代表者、実施者、協力者として関与」76.5%、「研究開発領域・プログラムのマネジメントに、総括やアドバイザー等として関与」27.9%であった。

3.結果

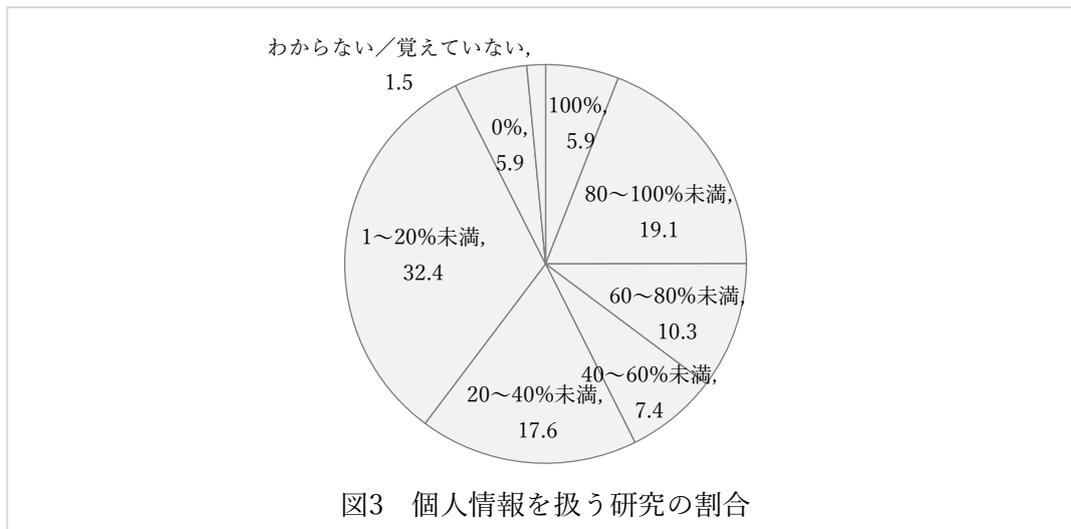
(1)取扱う研究データの種類・性質等

まず、扱った経験のある研究データの種類についての回答結果（複数回答可）を図1に示す。「自身が調査や実験等により取得したデータ」88.2%が最も多く、「自身が調査会社へ調査を依頼して取得したデータ」60.3%、「企業や自治体等の他機関が保有する公開データ」57.4%、「パネルや統計等の公開データ」55.9%、「研究者が公開しているデータ」51.5%と続いた。非公開・限定公開のデータは公開データと比べるとポイントは下がるが、「企業や自治体等の他機関が保有する非公開データ」30.9%、「研究者が限定した人のみに公開しているデータ」29.4%であった。

次に、扱った経験のある研究データの性質についての回答結果（複数回答可）を図2に示す。「観察・観測データ（社会現象の観察、自然現象の観測によって得られたデータ：社会調査、インタビュー、地震観測など）」89.7%が最も多く、「派生データ・編集されたデータ（加工・結合されたデータ：テキストマイニングなど）」36.8%、「実験データ（実験によって得られたデータ：DNAシーケンスなど）」35.3%と続いた。



続いて、普段の研究のうち個人情報データを扱う研究の割合について質問した回答結果を図3に示す。「100%」5.9%、「80%～100%未満」19.1%、「60%～80%未満」10.3%、「40%～60%未満」7.4%、「20%～40%未満」17.6%、「1%～20%未満」32.4%、「0%」5.9%、「わからない・覚えていない」1.5%であった。



さらに、研究倫理審査に関連した質問もした。直近の研究倫理審査でのデータに関わる審査についてあてはまるものを、複数回答を許す形式で尋ねたところ、「データの取得について、研究倫理審査の中で審査された」58.8%、「データの管理について、研究倫理審査の中で審査された」55.9%、「データの利活用について、研究倫理審査の中で審査された」54.4%となり、データの取得、管理、利活用について大きな差は見られなかった。また、「研究倫理審査は受けたが、データについて議論・意見はなかった」は13.2%であった。なお、研究倫理審査経験者(n=52)には申請方法についても尋ねているが、「自分で申請した」71.2%、「第三者が代理で申請した」26.9%であった。

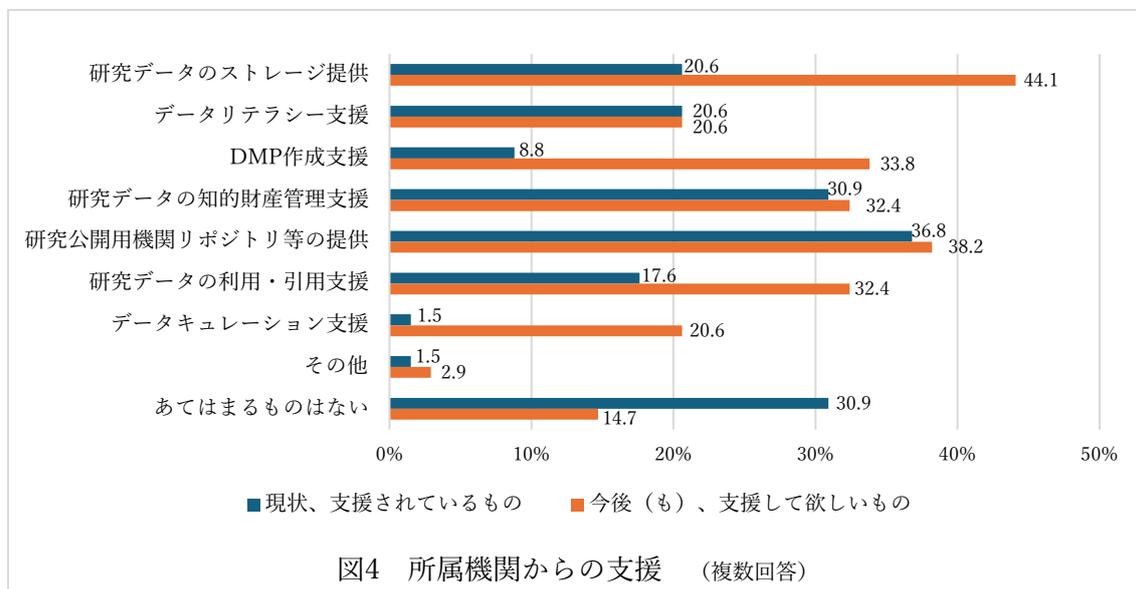
(2) 研究データに関する所属機関との関係

まず、所属機関の研究データポリシーへの理解度について、最もあてはまるものを尋ねたところ、「内容を理解していて、守っている」が82.4%と圧倒的に多く、「所属機関にポリシー等が存在しているかわからない」7.4%、「存在は知っているが、内容は理解していない」4.4%であった。なお、「所属機関にポリシー等は存在しない」4.4%は、所属機関を企業とした回答であった。

次に、所属機関からの支援について、①現状、支援されているもの、②今後(も)、支援してほしいものを質問した回答結果(複数回答可)を図4に示す。

①現状支援されているものの上位3位は、「研究公開用機関リポジトリ等の提供」36.8%、「研究データの知的財産管理支援」30.9%、「あてはまるものはない」30.9%であった。②今後(も)、支援してほしいものの上位3位は、「研究データのストレージ提供」44.1%、「研究公開用機関リポジトリ等の提供」38.2%、「DMP(データマネジメントプラン)作成支援」33.8%であった。また、①現状に対して②今後の割合が多く、ポイント差が大きかったものは、「研究データのストレージ提供」、「DMP(データマネジメントプラン)作成支援」、「研究データの利用・引用支援」、「データキュレーション支援」であった。この設問は、大学等研究機関を対象とした JPCORE/AXIES 調査で用いられていた研究者からの要望が高いと

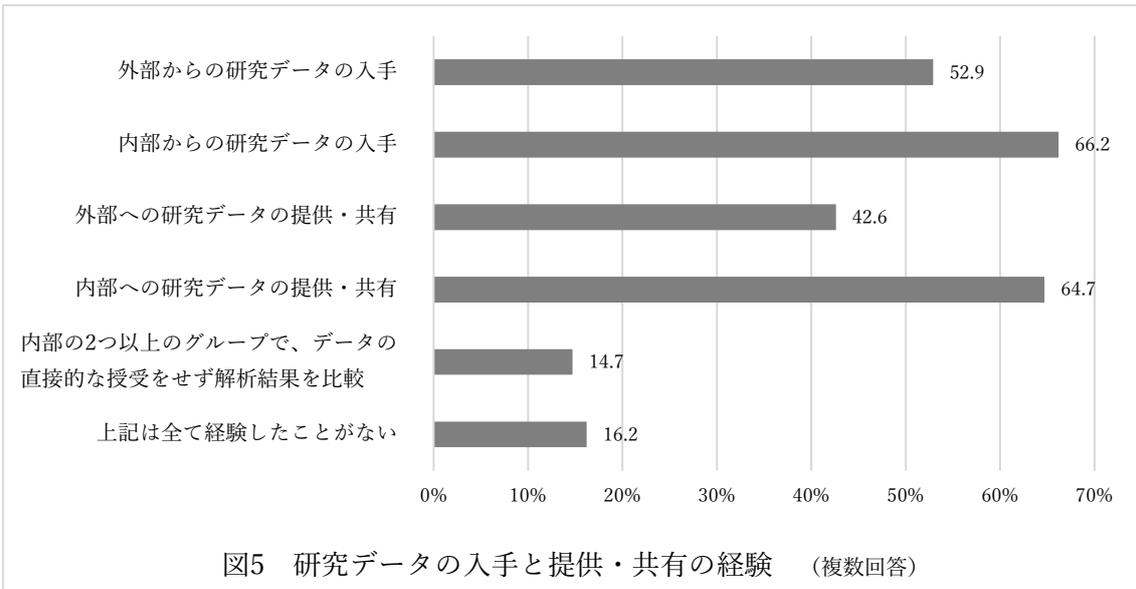
思われるサービスの項目を利用したが⁸、同調査では順位としては下位にあった「DMP（データマネジメントプラン）作成支援」が、本調査においては支援して欲しいものの上位となった。



(3) 研究データの入手と提供・共有

研究データの入手と提供・共有についての項目では、RISTEX や科研費等の複数のメンバーで行う場合を「プロジェクト（研究課題）」と定義した。同じく、そのメンバーを「内部」、メンバー以外を「外部」と定義して質問した。

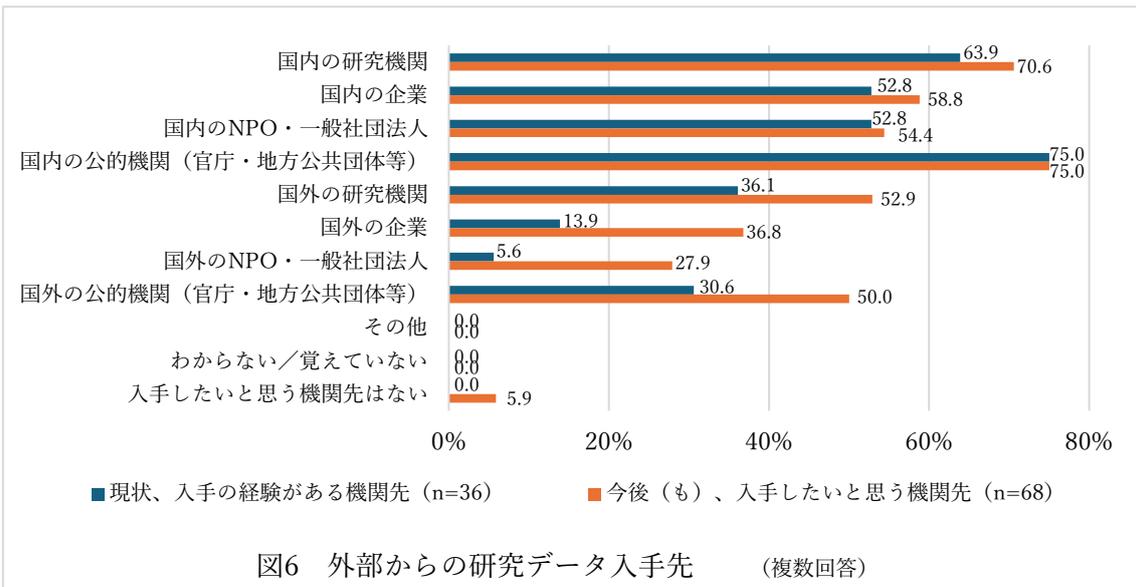
まず、プロジェクト内外でのデータの入手と提供・共有の経験についての回答結果（複数回答可）を図5に示す。「外部からの研究データの入手」52.9%、「外部への研究データの提供・共有」42.6%と、外部からの入手に対して外部への提供・共有が10.3ポイント低かった。内部の場合は「内部からの研究データの入手」66.2%、「内部への研究データの提供・共有」64.7%とほとんど差がなかった。また、「内部の2つ以上のグループで、データの直接的な授受をせず解析結果を比較」14.7%、「上記は全て経験したことがない」16.2%であった。



(3)-1.研究データの入手

研究データの入手経験がある場合は、データの外部からの入手先、入手した研究データの加工段階、入手時の対応、問題点の4項目について質問した。まず、データの外部からの入手先について、①現状、入手の経験がある機関先 (n=36)、②今後 (も)、入手したいと思う機関先 (n=68) についての回答結果 (複数回答可) を図6に示す。

①入手の経験がある機関先の上位3件については、「国内の公的機関 (官庁・地方公共団体等)」75.0%、「国内の研究機関」63.9%、「国内の企業」と「国内のNPO・一般社団法人」が52.8%で、国内の機関先が占めた。②今後 (も)、入手したいと思う機関先は、上位については①現状と大きな差はなかった。①現状と②今後を比較すると、国外の機関先について①現状に対して②今後の割合が多く、ポイント差が大きかった。

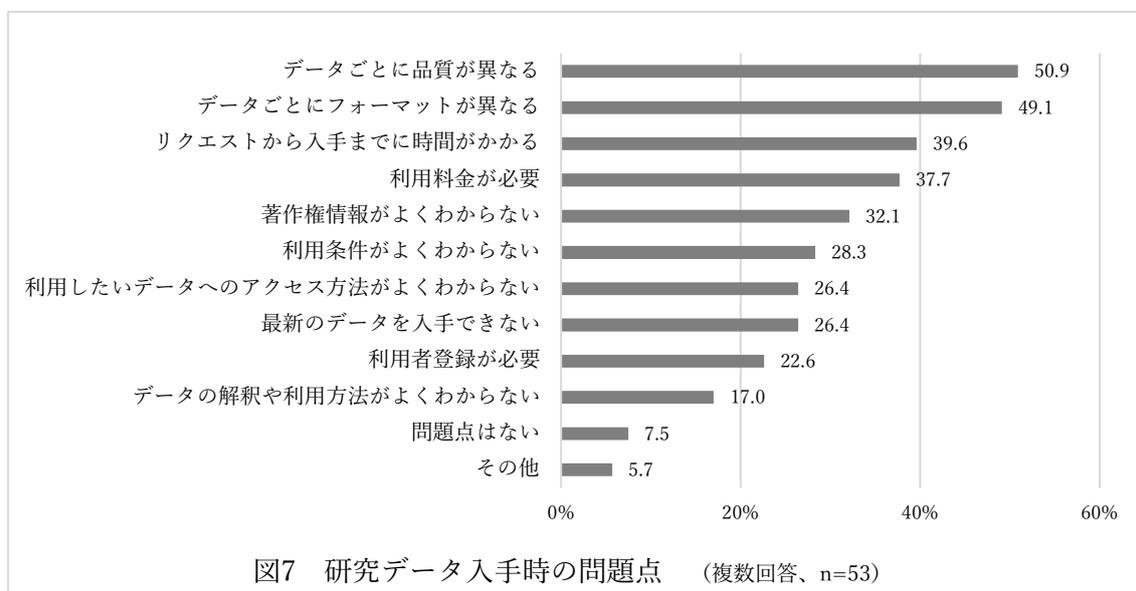


次に、入手する研究データの加工段階についての回答結果 (複数回答可) は、「生データ

（匿名化等の処理を行っていないデータなど）」52.8%、「分析可能なデータセット（匿名化等の個人の特定可能性を下げる一定の処理を行ったデータなど）」73.6%、「要約データ（個人の特定ができないように統計処理されたデータなど）」64.2%、「論文等に記載された裏づけデータ」39.6%であった。

続いて、研究データ入手時の対応として経験があるものについての回答結果（複数回答可）は、「機関間で文書の取り交わし（契約）を行う」66.0%、「個人間で文書の取り交わし（契約）を行う」20.8%、「文書の取り交わしはせず、メールや口頭で利用を約束する」41.5%、「何もしない」15.1%であった。

さらに、研究データ入手時の問題点についても尋ねたが、その回答結果（複数回答可）は図7に示す。「データごとに品質が異なる」50.9%、「データごとにフォーマットが異なる」49.1%が高い結果となった。この設問は、NISTEP 調査において用いられているデータ入手の障壁の項目を利用したが、データの品質とフォーマットの問題は NISTEP 調査でも例年上位にある項目に該当する⁹。

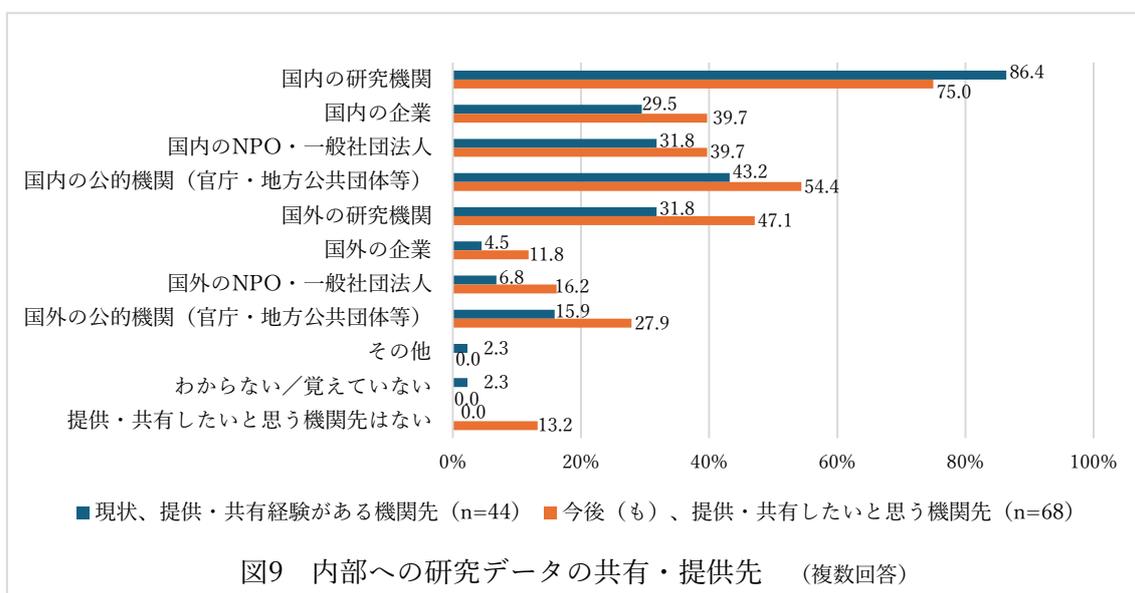
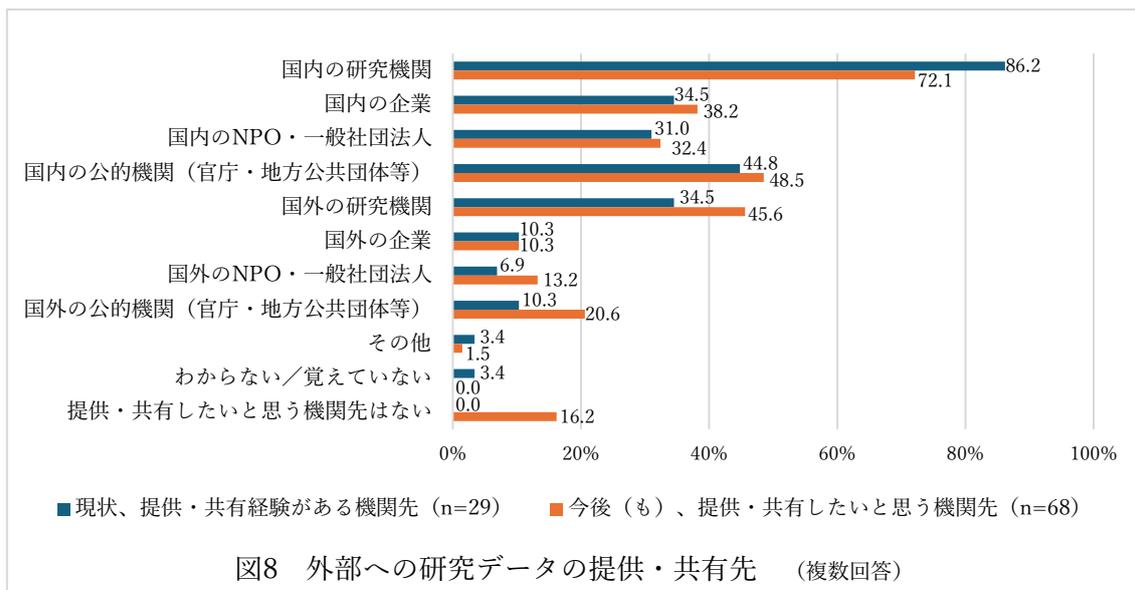


(3)-2.研究データの提供・共有

研究データの提供・共有経験がある場合は、共有・提供先を外部と内部に分けて、データの提供・共有先、提供・共有する研究データの加工段階、提供・共有時の対応、懸念点の4項目について質問した。

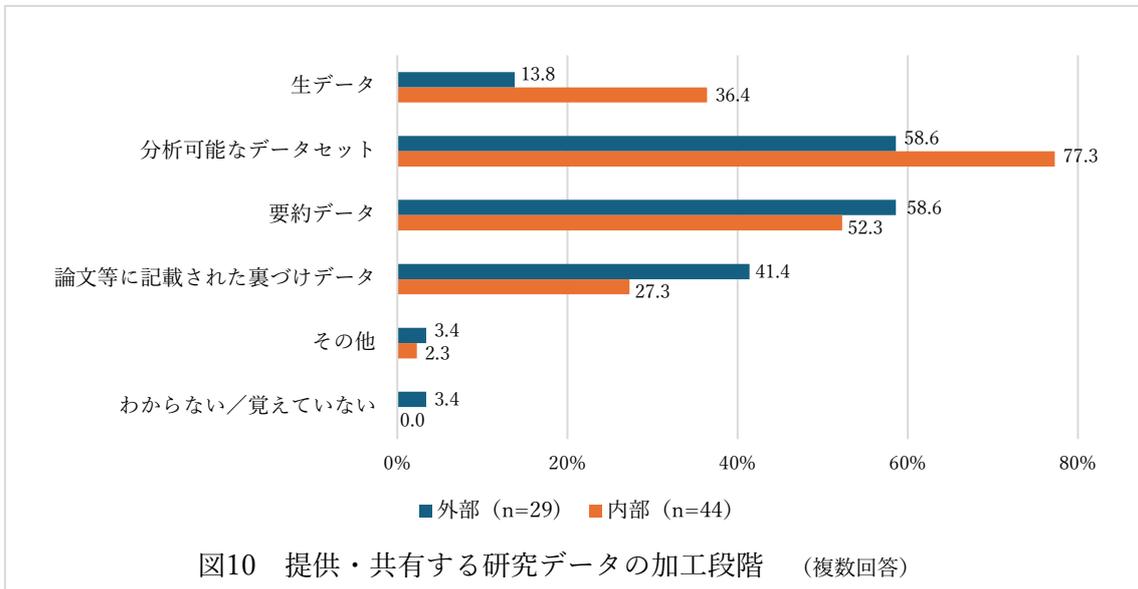
まず、外部への研究データの提供・共有先について、①現状、提供・共有の経験がある機関先 (n=29)、②今後(も)、提供・共有したいと思う機関先 (n=68) についての回答結果（複数回答可）を図8に示す。①提供・共有の経験がある機関先の上位3位は、「国内の研究機関」86.2%、「国内の公的機関（官庁・地方公共団体等）」44.8%、「国内の企業」と「国外の研究機関」が34.5%であった。②今後(も)、提供・共有したいと思う機関先は、「国内の研究機関」72.1%、「国内の公的機関（官庁・地方公共団体等）」48.5%、「国外の研究機関」45.6%であった。次に、内部への研究データの提供・共有先について、①現状、提供・

共有の経験がある機関先 (n=44)、②今後 (も)、提供・共有したいと思う機関先 (n=68) についての回答結果 (複数回答可) を図9に示す。①提供・共有の経験がある機関先の上位3位は、「国内の研究機関」86.4%、「国内の公的機関 (官庁・地方公共団体等)」43.2%、「国内のNPO・一般社団法人」と「国外の研究機関」が31.8%であった。②今後 (も)、提供・共有したいと思う機関先は、「国内の研究機関」75.0%、「国内の公的機関 (官庁・地方公共団体等)」54.4%、「国外の研究機関」47.1%であった。このように、研究データの提供・共有先については、外部と内部で①現状と②今後の上位について顕著な差はみられなかった。

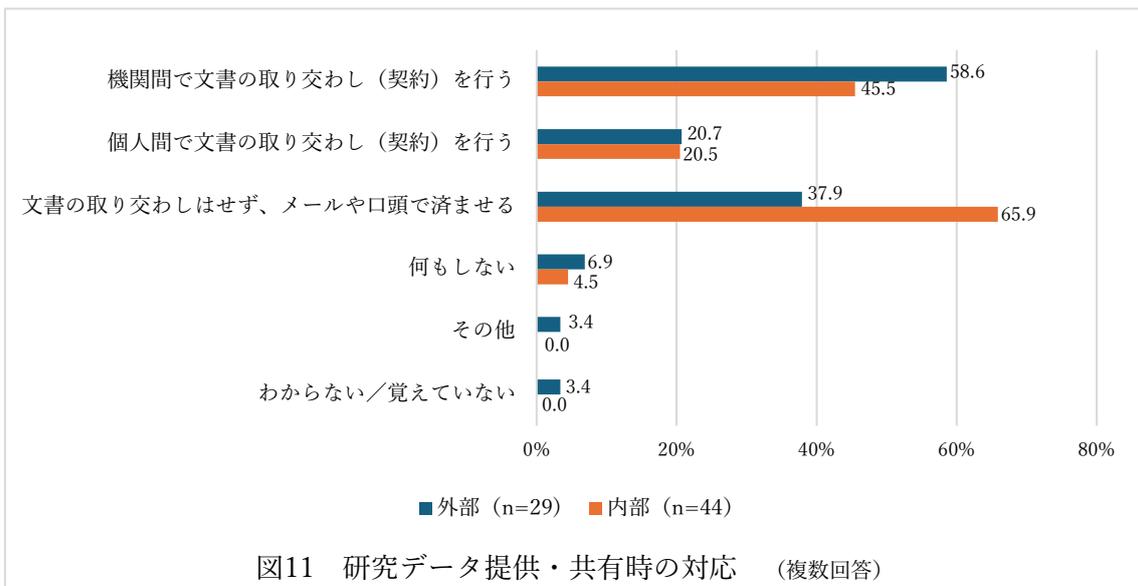


次に、提供・共有する研究データの加工段階¹⁰については、外部と内部の回答結果 (複数回答可) をまとめて図10に示す。「分析可能なデータセット (匿名化等の個人の特定可能性を下げる一定の処理を行ったデータなど)」が外部58.6%と内部77.3%と高く、次いで、「要約データ (個人の特定ができないように統計処理されたデータなど)」が外部58.6%と内部

52.3%となった。「生データ（匿名化等の処理を行っていないデータなど）」外部 13.8%と内部 36.4%と 22.6 ポイント差、「分析可能なデータセット」も 18.7 ポイント差があり、提供・共有先が外部と内部で差がみられた。

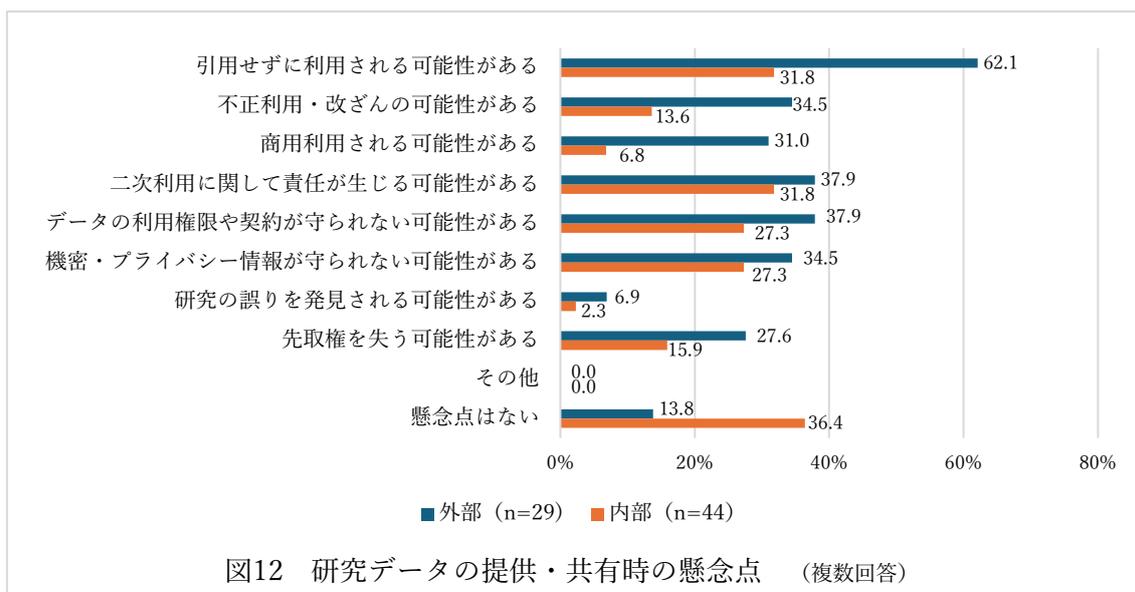


続いて、研究データの提供・共有時の対応として経験があるものについて、外部と内部の回答結果（複数回答可）をまとめて図 11 に示す。外部への提供・共有時は、「機関間で文書の取り交わし（契約）を行う」58.6%が最も高く、次いで「文書の取り交わしはせず、メールや口頭で利用を約束する」37.9%であった。一方、内部への提供・共有時は、「文書の取り交わしはせず、メールや口頭で利用を約束する」65.9%が最も高く、次いで「機関間で文書の取り交わし（契約）を行う」45.5%であった。「文書の取り交わしはせず、メールや口頭で利用を約束する」については外部と内部で 28 ポイント差があった。



さらに、研究データの提供・共有時の懸念点について、外部と内部の回答結果（複数回答

可)をまとめて図12に示す。外部への提供・共有時は、「引用せずに利用される可能性がある」62.1%が最も高くなった。この設問も NISTEP 調査において用いられている項目を利用したが、引用の問題は例年上位にある項目に該当する¹¹。次いで、「二次利用に関して責任が生じる可能性がある」と「データの利用権限や契約が守られない可能性がある」が37.9%、「不正利用・改ざんの可能性がある」と「機密・プライバシー情報が守られない可能性がある」が34.5%と高い結果となった。一方、内部への提供・共有時は「懸念点はない」が36.4%と最も高くなった。



最後に、研究データ管理や調査に関する意見・感想の自由回答については、特にない等を除き14件の記述があった。本調査に関わる意見2件以外は研究データ管理に関わるものであり、データフォーマットの統一、データ管理のリテラシーの向上、データ公開のノウハウやインフラ等の必要性が挙げられた他、研究機関の体制の現状、オープンアクセス対応等に対する指摘があった。また、研究開発プロジェクト終了後を含めた長期的な研究データ管理に関する意見が複数あった。

4.おわりに

世界的に研究データの公開や共有に必要な支援は不足し、適切な支援体制を構築することが喫緊の課題であること、また、支援に際しては分野の違いを考慮した繊細なアプローチが必要であること等が指摘されている¹²。本調査から見てきた社会技術研究開発における研究データの特徴や研究者の要望を参考にしながら、RISTEXとしてよりよい研究データ管理に向けた取り組みを進めたい。

注

- ¹ 研究活動における研究データの収集、整理、保管、文書化等の取り扱いを指す。
- ² 令和 5（2023）年度から試行的に取り組みを開始している。データマネジメントについて専門的助言等を行うデータマネジメント・アドバイザーボードを設置し、データマネジメントプラン（DMP）の改訂や事例検討等を行う他、「ジャパン・オープンサイエンス・サミット 2024（JOSS2024）」においてセッションを開催する等、研究データ管理に関心を持つ広い関係者と共に課題を検討する機会も設けている。なお本調査は、JOSS2024 の RISTEX セッション「〔総合知〕活用における研究データマネジメント～現状と課題」において速報値を報告したものである。
- ³ 池内有為、林和弘「研究データ公開と研究データ管理に関する実態調査 2022：日本におけるオープンサイエンスの現状」NISTEP RESEARCH MATERIAL、No.335、文部科学省科学技術・学術政策研究所、DOI: <https://doi.org/10.15108/rm335>。また、同調査を主に人文・社会科学の観点から分析している池内有為「人文学・社会科学分野におけるデータ管理と利活用に資する専門人材の育成と課題」（人文学・社会科学特別委員会（第 20 回）配付資料 令和 5 年 12 月 22 日）https://www.mext.go.jp/content/20231222-mxt_sinkou01-000033251_02r.pdf も参照した。
- ⁴ 池内有為、林和弘「日本の研究機関における研究データ管理（RDM）の実践状況－オープンサイエンスの実現に向けた課題と展望」STI Horizon、Vol.8 No.1、2022、DOI: <https://doi.org/10.15108/stih.00287>。
- ⁵ SDGs の達成に向けた共創的研究開発プログラム（シナリオ創出フェーズ・ソリューション創出フェーズ）、SDGs の達成に向けた共創的研究開発プログラム（社会的孤立・孤独の予防と多様な社会的ネットワークの構築）、SDGs の達成に向けた共創的研究開発プログラム（情報社会における社会的側面からのトラスト形成）、科学技術の倫理的・法制度的・社会的課題（ELSI）への包括的実践研究開発プログラム、科学技術イノベーション政策のための科学 研究開発プログラムである。
- ⁶ 工学、天文学、化学、物理学、地球科学、数学、計算機科学（コンピュータサイエンス）、農学、生物科学、医学、心理学、社会科学、人文学の 13 分野であり、これは注 3「研究データ公開と研究データ管理に関する実態調査 2022」を参照した。なお、NISTEP の同調査の結果は工学の割合が最も多くなっている。
- ⁷ トランスディシプリナリー（TD）研究は、「ある社会問題の解決に向けて研究者とその課題に関わる人びと（ステークホルダー）が、研究開発のデザインから実施、そして研究開発で生み出した解決策の社会実装までを一緒に進めるようなスタイルの研究」として設問に示した。本定義は、RISTEX が実施した「トランスディシプリナリー（TD）研究の評価実践に向けた 調査・試行・検討」（https://www.jst.go.jp/ristex/internal_research/files/r5fesurvey_fullreport.pdf）における調査で用いた定義を利用した。
- ⁸ 注 4「日本の研究機関における研究データ管理（RDM）の実践状況」54 頁、「研究データ

管理サービスを展開するにあたり、自機関の研究者から要望が高いと思われるサービスをお聞かせください」。

- ⁹ 注3「研究データ公開と研究データ管理に関する実態調査 2022」65頁、「Q8 公開データを入手した際に、問題だと感じたことがある項目」。経年変化については、同24頁の「データ入手の障壁の経年変化」を参照。
- ¹⁰ データの加工段階については、Lipton V. J. Open Scientific Data -Why Choosing and Reusing the RIGHT DATA Matters. Intech Open, 2020、DOI: <https://doi.org/10.5772/intechopen.87201>、厚生労働省保険局医療介護連携政策課 保険システム高度化推進室「第1回 NDBオープンデータ 【解説編】」（平成28年10月）<https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000141549.pdf>を参考にした。
- ¹¹ 注3「研究データ公開と研究データ管理に関する実態調査 2022」74頁、「Q23 カレントデータを公開しようとする場合、次の点ほどの程度問題となりますか」の項目を、本調査では提供・共有時の設問として利用した。経年変化については同51頁の「データ公開に対する懸念」を参照。
- ¹² 注3「研究データ公開と研究データ管理に関する実態調査 2022」60頁。世界的な動向については、Digital Science、Springer Nature、Figshare が実施した“The State of Open Data 2023”<https://www.digital-science.com/news/state-of-open-data-report-2023/>を参照。