

資料 1-1

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
(第 11 回)
令和元年 8 月 20 日(火)

2019/08/20

研究データ基盤整備と国際展開ワーキング・グループ報告書 「研究データ基盤整備と国際展開に関する戦略」

1. 背景と目的

1-1. オープンサイエンスに関する国際的動向

研究開発の過程で生まれるデータ¹（以下、「研究データ」という。）は、我が国及び世界にとって重要な知的財産である。したがって、それら研究データを有効に利活用することにより、知の発展と結合を促し、優れた研究成果やイノベーションの創出へとつなげることが重要である。研究データの利活用に関しては、国際的には、知識をオープンにし、研究の加速化や新たな知識の創造などを促すオープンサイエンスの動きが活発化している。しかしその一方で、学術論文に関するビジネス環境においては世界的な出版社による寡占化が進行しており、さらに研究データについてもこれらの出版社や IT・データ関係企業等がビジネスとして焦点を当てつつある。これに対して、ドイツ、スウェーデン、フランス等では学術出版社との雑誌購読契約の見直しをする等の動きがある他、欧州では、域内の研究データを適切に管理・共有するための研究データ基盤として **European Open Science Cloud (EOSC)** の構築が進められている。

2013 年に英国で開催された G8 科学大臣及びアカデミー会長会合では、その共同声明において、論文のオープンアクセス化に加えて、研究データのオープン化についても言及された。また、2016 年 G7 茨城・つくば科学技術大臣会合においてオープンサイエンスが主要な議題として取り上げられ、EU と日本が共同事務局を務めるワーキング・グループが設置された。このように、多国間枠組みでのより具体的な検討も進んでいるところである。

1-2. 日本政府における取り組み

このようなオープンサイエンスに関わる国際的な動向に鑑み、公的資金²による研究データの管理及び利活用について、内閣府は、2015 年 3 月に「国際動向を踏まえたオープンサイエンスに関する検討会」での報告書「我が国におけるオープンサイエンス推進のあり方について」において基本的な考え方をとりまとめるとともに、第 5 期科学技術基本計画においてオープンサイエンスへの対応の重要性を明記した。第 5 期科学技術基本計画では、サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会（Society）として、**Society 5.0** を提唱しており、オープンサイエンスへの対応はこれを支える重要な要素である。

「第 5 期科学技術基本計画」におけるオープンサイエンスに関する記述（抜粋）

国は、資金配分機関、大学等の研究機関、研究者等と連携し、オープンサイエン

¹ 研究開発の過程で生み出されるデータ。電磁的な形態により管理が可能なものを意味し、サンプル等の実体物は含まない。研究データの範囲については、参考 1 に示す。

² 公的資金とは、国又は資金配分機関（我が国の公的資金を配分する機関）から大学、研究開発法人、企業等に対して交付、補助又は委託する全ての経費を対象とする。

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。なお、研究分野によって研究データの保存と共有の方法に違いがあることを認識するとともに、国益等を意識したオープン・アンド・クローズ戦略及び知的財産の実施等に留意することが重要である。

また、日本政府は、科学技術基本計画を具現化する年次の戦略である「統合イノベーション戦略 2018」において、知の源泉を支える3つの基盤の一つとして、オープンサイエンスの推進の一節を設けて、以下のように規定している。

「統合イノベーション戦略 2018」におけるオープンサイエンスに関する記述（抜粋）

今後の方向性及び具体的に講ずる主要施策

研究データを我が国のデータインフラから公開できるよう、主として機関リポジトリを対象としたシステム開発や、国際認証基準等を参考にしたリポジトリの整備・運用を進める。

その上で、研究分野の特性等を踏まえたオープン・アンド・クローズ戦略を考慮したデータポリシーやデータマネジメントプランの策定を促進し、これらに基づく研究データの管理・公開等を促進するとともに、公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する。

さらに、これらの取組が大学・国研等で適切に行われるよう、研究データの管理や公開・共有に従事する研究者等の意識向上や基礎的な知識の習得のための取組や、研究者や大学・国研等における現状、取組等についての調査・分析を行い、研究者等の意識向上等に資する方策を検討する。

i) リポジトリの整備及び展開【科技、文】

・文部科学省が主体となり、機関リポジトリを活用しクラウド上で共同利用できる研究データの管理・公開・検索を促進するシステムを開発し、2020年度に運用開始

・国際認証基準等に基づくリポジトリの整備・運用のガイドライン（公開データの検索可能化、諸外国の研究データ基盤との相互運用性等を含む。）を内閣府（科技）が策定し、大学・国研等にガイドラインの適用を推奨

ii) 研究データの管理・利活用についての方針・計画の策定等【内閣官房、科技、

食品、総、文、厚、農、経、国、環、防】

・内閣府（科技）は、国研におけるデータポリシーの策定を促進するためのガイドラインを2018年6月までに策定

・国研は、研究分野の特性、国際的環境、産業育成等に配慮し、必要に応じてオープン・アンド・クローズ戦略を取り入れ、データポリシーを策定

・競争的研究費制度の目的、対象等を踏まえ、大学・国研・企業等の研究実施者がデータマネジメントプラン等のデータ管理を適切に行う仕組みを、各府省・研究資金配分機関が所管の競争的研究費制度に導入^{注)}（ガイドライン策定や公募要領改訂等）

・データポリシー・データマネジメントプランに基づく公的資金による研究データの管理・公開等を促進し、公的資金による研究成果としての研究データについてはデータインフラを通して機械判読可能化を促進

注) 導入府省・研究資金配分機関は、2018年度当初時点で4であるが、2021年度予算における公募までに14府省・機関全てでの導入を目指す（制度内で特定のプログラム・事業等で導入した場合を含む。）。

iii) 人材の育成及び研究データ利活用の実態把握【科技、文】

・2017年度に開発・公開された基礎的な能力開発プログラムについて、専門性を高めた拡張版の開発・公開及び関係機関における受講の促進

・大学・国研等のデータポリシー等作成状況、リポジトリ整備状況、研究データ利活用の優良事例、研究者の公的資金による研究データの公開状況や利活用上の障壁・促進要因等の調査を2019年度から実施

これらを踏まえ、内閣府「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」は、2018年6月に「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」、2019年3月に「研究データリポジトリ整備・運用ガイドライン」を策定した。

1-3. 研究データ基盤整備と国際展開ワーキング・グループの設置

以上のような背景から、我が国においても、研究データ基盤を構築し、研究データの適切な管理及び利活用を促すことにより、我が国の研究コミュニティの自律性を確保しつつ、優れた研究成果を生み出し、イノベーションにつなげる必要がある。そこで、そのための国としての戦略と具体的な施策を明確にすることを目的として、2018年12月、上記「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」の下で「研究データ基盤整備と国際展開ワーキング・グループ（WG）」の設置が承認され、2019年2月より、主に以下の事項に関する検討が行われた。本報告書ではそれらの検討内容について報告する。

- 我が国の研究データを適切に管理・共有・公開・検索³するための研究データ基盤シ

³ 本報告書では、特記なき場合には、管理・共有・公開・検索の一連の行為を「管理及び

システムの構築や政府全体の体制に関する検討

- 国レベルでの研究データ戦略やデータポリシー等に関する検討
- 国際的な対応のあり方の検討
- 地球環境測定系、生命科学系、材料科学系、工学系等の各研究分野における最先端研究者の研究データマネジメントに関するケーススタディ

2. 公的資金を活用して生み出された研究データの管理・利活用の基本方針

2-1. 基本的な考え方

本 WG の検討結果として、公的資金を活用して生み出された研究データの適切な管理及び利活用のための、研究データ基盤の整備とこれに関連する制度及び国際展開に関する戦略として、基本的な考え方を示す。

国は、本報告書を踏まえ、以下の諸点を考慮し、国全体の研究データ管理及び利活用に関する基本方針（ナショナル・データ・ポリシー）を示すこととする。

- 研究データは、我が国と世界にとって重要な知的財産であり、その利活用を進めることは、さまざまな知の発展と結合による優れた研究成果やイノベーションの創出につながるものである。
- 世界的なオープンサイエンスの潮流を尊重しつつ、我が国の自律性を確保し、オープン・アンド・クローズ戦略を考慮する必要がある。
- 研究データの管理及び利活用の方針や具体的な方策については、一般に、研究成果の最大化や波及効果の誘起、さらには研究成果の利活用を通じた新たな研究成果の創出を期待する観点から立案・実施される。その際、組織におけるデータポリシーの策定や、資金配分機関等におけるデータ・マネジメント・プラン（DMP）の導入にあたっては、個々の組織（大学、研究開発法人、民間企業など）や競争的研究費制度の特性（産業化を目指すものか、学術目的のものかなど）に配慮する必要がある。
- 企業、大学、研究機関等の研究活動にとって、新製品の研究開発情報や研究の新規性を有する研究データは死活的に重要である。また研究データは、国の安全保障を確保し、我が国の産業競争力や科学技術・学術上の優位性を確保するために重要な情報を含む可能性がある。このため、我が国としては、個人情報、企業の秘密情報、研究の新規性、我が国の安全保障、我が国の産業競争力、我が国の科学技術・学術上の優位性等の観点から留意すべき研究データは非公開とし、その他の研究データについては原則公開とする。なお、先進的な研究データ・マネジメントの試行や国内外の優良事例の収集等を通じて、研究データの類型化や望ましい研究データの管理及び利活用のあり方に関して具体的な検討を進めることが必要である。
- 研究データの管理及び利活用にあたっては、公開・非公開に関わらず、十分なセキュリティが確保される必要がある。
- 外国企業、政府等の外的な状況の変化によって、我が国から公的資金による研究デー

利活用」としている。なお、これらの用語の国際的な通用性については、今後検討を要する。

タにアクセスや利活用が出来ない状態になることを避けるとともに、我が国の安全保障、産業競争力や科学技術・学術的な優位性の観点から重要な情報が不用意に流出することが無いように措置する必要がある⁴。

- 研究データの管理及び利活用にあたっては、公開・非公開に関わらず、関係諸法令⁵に従うとともに、データの取り扱いに関する条約等の国際約束等⁶や、データ管理の原則である FAIR 原則⁷等の国際的な規則や慣行等との整合性に十分留意する。
- 論文や特許等の直接的な研究成果のみならず、研究の過程で生まれた一定の体系性を有するデータの管理及び利活用を行うことは、特に基礎的な研究の価値を高める上で有意義なことである。また、このような取組は、必ずしも直接的な成果とならなくても、研究者に対して、失敗を許容し、様々な挑戦を促すことにつながるものである。

本 WG の審議状況を踏まえ、統合イノベーション戦略 2019（令和元年 6 月 21 日 閣議決定）において、以下のように関連する記述がなされている。

統合イノベーション戦略 2019（2）研究データ基盤の整備・国際展開（抜粋）

目指すべき将来像

- ・ 国益や研究分野の特性等を踏まえて、オープン・アンド・クローズ戦略を考慮し、サイバー空間上での研究データの保存・管理に取り組み、諸外国の研究データ基盤とも連携して巨大な「知の源泉」を構築し、あらゆる者が研究成果を幅広く活用
- ・ その結果、所属機関、専門分野、国境を越えた新たな協働による知の創出が加速

< 中略 >

目標達成に向けた施策・対応策

< 研究データ基盤及びリポジトリの整備 >

- 国の各機関において、研究者が信頼して研究データを保存・管理できるためのリポジトリの整備を推奨する。【全府省庁】
- 研究データ基盤システムは、文部科学省が主体となり、2020 年度内の本格運用に向けて開発を進める。【文】
- 公的資金による研究活動により生み出された研究データの利活用とそれによる先進的な研究マネジメント支援を促進するため、ムーンショット型研究開発

⁴ アクセス制限等の物理的な制限や情報管理に関する法的な制限の双方が考えられる。日本に所在するサーバー等に研究データを保管することが望ましい。

⁵ 個人情報の保護に関する法律、輸出貿易管理令、不正競争防止法（限定提供データ）など。研究データ利活用に関する法令等を参考 2 に示す。

⁶ GDPR（EU 一般データ保護規則：General Data Protection Regulation）、ロンドン・ガイドラインなど

⁷ データ共有の基準としての FAIR 原則 <https://doi.org/10.18908/a.2018041901>

制度において、先行的に研究データ基盤システムの活用を図るなど、先進的なデータマネジメントを推進する。【科技、文、経】

- ドメイン語彙策定への関係機関の参画等を通じ、研究データ基盤システムと分野間データ連携基盤との連携を図る。【科技、文】

＜研究データの管理・利活用についての方針・計画の策定等＞

- 研究データの管理・公開・検索のためのグランドデザインを踏まえ、基盤構築や体制整備等に関係するガイドラインを策定するとともに、データマネジメントプランで要求すべき事項等を取りまとめる。【科技】
- 「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」に基づき、国研においてオープン・アンド・クローズ戦略を踏まえたデータポリシーの策定を推進する。【全府省庁】
- 国研、大学等におけるデータリポジトリの整備・運用に当たっては、「研究データリポジトリ整備・運用ガイドライン」等を参考に、研究データの保存・公開等を適切に進める。【全府省庁】
- 各府省、研究資金配分機関が所管する競争的研究費制度において、データマネジメントプランの策定を研究実施者に求める仕組みの導入を進める。【全府省庁】
- 我が国の研究活動の自律性の確保とオープンサイエンスの推進のため、G7の枠組み等を活用し2019年6月に開催したワークショップでの結果も踏まえ、EU等の対応するシステムとの相互運用性を確保する等の提携に向けた検討を行う。また、国際協力の下、論文や研究データの寡占による研究開発活動への弊害を防ぐための対策を検討する。【科技】
- 内閣府（科技）は、学術の発展や研究成果の再現性を高めること等に資する情報の質の管理やコミュニティ間の共有等について、学問分野の特性や国際的動向を踏まえつつ、我が国の方向性に関するアカデミアの立場からの検討との連携を図る。【科技】

注）研究データの管理・利活用の方針については国研が2020年度末までに策定、計画については競争的研究費による研究実施者が策定することを要請する制度を2021年度予算における公募までに、各府省・研究資金配分機関において導入。

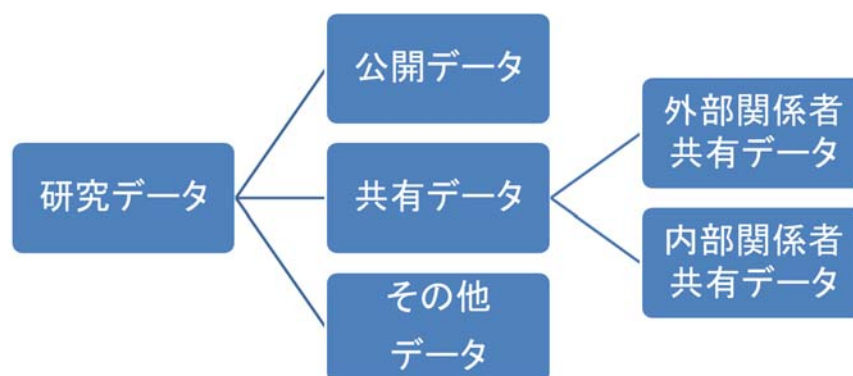
2-2. 研究データの管理及び利活用に関する方策

2-2-1. 研究データの区分

研究データについては、その特性や帰属先など様々な視点からの区分方法があるが、ここでは、利活用の視点から、図1のように区分して検討する。

研究データの管理にあたっては、まずは、研究者が実験や観測等の研究活動で生まれるデータを体系的に保存することが重要である。また、共同研究等のために内部の関係者や外部の関係者との間でデータの共有を進めることが望ましい。さらに、広く第三者による利活用

が見込まれるデータについては積極的に公開していくことが望ましい。その際には、研究データの横断的かつ効率的な検索を可能にするために、国際的通用性のある識別子やメタデータの付与が行われることが適当である⁸。



研究データ：研究開発の過程で生み出されるデータ

公開データ：論文や特許公報に掲載されたデータ、その他一般に公開されるデータ

共有データ：限定された関係者での利活用を目的として共有されるデータ

外部関係者共有データ：機関やプロジェクトの外部の関係者と共有されるデータ

内部関係者共有データ：機関やプロジェクトの内部の関係者と共有されるデータ

その他データ：研究データのうち、公開も共有もされないもの⁹

図1 利活用の視点からの研究データの区分

2-2-2. 組織における研究データ管理及び利活用とデータポリシー

研究データは、研究者の所属する大学、研究機関等が組織としてデータポリシーを策定し、その下で適切に管理される必要がある。内閣府は、統合イノベーション戦略2018に基づき、2018年6月に「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」を策定しており、2019年6月までに、7つの研究開発法人がこのガイドラインに従ってデータポリシーの策定を完了している。引き続き、その他の研究開発法人や大学等の研究機関においてもデータポリシーの策定を進めていくことが必要である。

また、公的資金のうち、競争的資金及び公募型資金等の機関の外部から受け入れる資金によって生み出された研究データについては、公的資金を受けた研究機関又は研究者が所属する機関（責任機関）の責任の下で管理することが一般的であり¹⁰、当面の考え方として責

⁸ 経済産業省では、本省及びNEDOの委託研究開発で得られた研究データの利活用を図るために、ガイドライン（「委託研究開発におけるデータマネジメントに関する運用ガイドライン」（経済産業省平成29年12月）。以下「経済産業省ガイドライン」という。）を策定するとともに、ナショナルデータカタログを公開している。

⁹ 一般的には、実験の試行等に伴うデータ、備忘録、研究ノート等が想定される。

¹⁰ 委託費についての知的財産権は本来委託者に帰属するが、バイドール制度（産業技術力強化法第17条）により、知的財産権について委託先に帰属されうる。また、交付金及び補

任機関のデータポリシーに従うことが適当である。競争的資金や公募型資金等については、それぞれの資金配分機関が、独自のデータポリシーを定めていることがあるが、双方のデータポリシーに相反が無く、研究者に過重な負担をかけることの無いようにすることが必要であり、国全体として整合性のある対応が望まれる。また、研究者においては、責任機関のデータポリシーと資金配分機関のデータポリシーの双方の遵守が求められる。なお、資金配分機関のデータポリシーにおいて、積極的な管理が求められていない場合¹¹においても、責任機関のデータポリシーの下で適切に管理されることが重要である。

2-2-3. 研究データ管理及び利活用のためのリポジトリの整備

研究データは、機関リポジトリ又は分野別リポジトリに保存及び管理されることが望ましい¹²。しかしながら、我が国における機関リポジトリの数は世界でも最多であるにも関わらず、研究データの登載が進んでおらず、積極的な利活用が望まれる。内閣府では、統合イノベーション戦略 2018 に基づき、2019 年 6 月に「研究データリポジトリ整備・運用ガイドライン」を策定し、各機関等におけるリポジトリの整備が進められている。さらに、統合イノベーション戦略 2019 では、これらの状況を踏まえ、国の各機関において、研究者が信頼して研究データを保存・管理できるためのリポジトリの整備を推奨している。これらのリポジトリについては、FAIR 原則やリポジトリの国際認証など、研究データの流通に関する国際的通用性を確保することが重要である。

また、国立情報学研究所（NII）が主体となって、機関リポジトリ等と連携して研究データを管理及び利活用するためのシステム（NII Research Data Cloud）の開発を進めている。

2-2-4. 競争的研究費制度と研究データの管理・利活用（データ・マネジメント・プラン（DMP）の導入など）

研究データの管理及び利活用の方針や具体的な方策については、一般に、研究成果の最大化や波及効果の誘起、さらには研究成果の利活用を通じた新たな研究成果の創出を期待する観点から立案・実施される。

DMP は、研究プロジェクト等の研究過程において、どのような種類のデータを、誰がど

助金については、その知的財産権は本来的に交付先に帰属することとなる。関係府省の研究助成プログラムでは、バイドール規定に基づく委託費及び補助金により執行されている場合が多く、そこで得られた成果物は受託者又は交付先が管理することが一般的である。研究データについては、産業技術力強化法の明示的な規定は無いが、同様に、その利活用のためには、責任機関において管理することが適切であると考えられる。経済産業省ガイドラインでは、「委託者が、研究データのうち、公共性が高く、広範な利活用を目的として国が管理する必要があるデータであって、その取得又は収集を目的として位置づけられ、国が指定したものに限り、国が管理することができる」としている。研究者個人に対する研究費の補助については、研究者が所属し研究費を執行管理する機関を責任機関としている。

¹¹ 経済産業省のガイドラインによる「非管理データ」などが該当する。

¹² リポジトリは、データインフラのうち、電子的な知的生産物の保存や発信を行うためのインターネット上のアーカイブシステム。国研、大学等が管理する機関リポジトリ、学会や特定分野の研究機関等が管理する分野別リポジトリ、及び分野・機関に限定されずデータ全般を対象とする汎用リポジトリがある。

のように取得し、どのように保存するかなどについて明記した計画書であり、研究者による研究データの適切な管理や効率的な研究進捗の把握を主な目的として、研究者や研究プロジェクトをマネジメントする者により作成される。2003年に米国衛生研究所（NIH）が、資金提供を受けた研究者に対してDMPの提出を義務化したことを端に発し、各国の資金配分機関においても同様の動きが広がりつつある。

日本政府においても、統合イノベーション戦略2018及び統合イノベーション戦略2019において、競争的研究費制度の目的や対象等を踏まえ、研究者がDMPを作成し、資金配分機関に提出することにより研究データを管理する仕組みを2021年度予算公募分までに導入することとしている。より詳細には、DMPにおいては、研究プロジェクト情報、研究データ情報、担当者情報、取得・収集情報等を記載することとする（別添の一般的なDMPの項目を参照）。内閣府や競争的研究費制度を所管する各省、資金配分機関等においては、一般的なDMP項目を参照しつつ、必要に応じて各事業の特性を踏まえて具体的な項目や記載内容等について検討する。また、内閣府は、競争的研究費制度におけるDMPの導入状況について、適時調査し、フォローアップすることとする。

資金配分機関等におけるDMPの導入にあたっては、個々の組織（大学、研究開発法人、民間企業など）や競争的研究費制度の特性（産業化を目指すものか、学術目的のものかなど）に配慮する必要がある。

2-2-5. 研究データの管理・利活用にあたっての留意事項等

① 個々の研究データの特性の違いに対する配慮

研究データの取り扱いにあたっては、学術データや産業データなどは、オープン・アンド・クローズ戦略などの観点においてそれぞれ性質が異なるため、大学、研究開発法人、企業等のデータの保有組織の特性に配慮した運用を行う必要がある。

産業競争力や科学技術・学術的な優位性を確保するためには、研究データを即時に公開することが適切で無い場合もありうる。公開による利活用の促進とのバランスを考慮しつつ、適切なエンバゴ（時限非公開）期間を設定することもあり得る。

② 組織外部からの資金によるプロジェクト等の終了後の研究データの取り扱い

競争的研究費制度の組織外部からの資金による研究プロジェクト等の終了後においても、取得された研究データが継続的に利活用されることを促進するとともに、不正な使用や、不用意な流出を防がなければならない。そのため、プロジェクト終了後の研究データの管理のあり方についても、研究プロジェクト等の実施段階から予め対応方法を検討しておく必要がある。

③ 研究公正のための研究データ管理

最近では、学術論文の査読にあたり、実験結果等のエビデンスとなる研究データの提供を求められる場合がある。また、組織及び研究者として研究公正の確保は重要な課題であり、実験データ等の適切な保存・管理が必要である。そのため、研究公正の確保のために、研究機関が所属研究者の実験データを体系的に保存、リストアップを行うなどの試みも進みつつある。また、画像認識技術を活用し、複製や改ざんなどを事前にチェックするソフトウェア

なども導入されつつある。

なお、これらの蓄積された研究データやリストを研究公正のみならず、積極的な利活用につなげていくことが今後の課題である。

3. 研究データ基盤と周辺環境の整備

3-1. 研究データ基盤システム

研究データを適切に管理及び活用するための研究データ基盤については、統合イノベーション戦略 2018 において「機関リポジトリを活用しクラウド上で共同利用できる研究データの管理・公開・検索を促進するシステムを開発し、2020 年度に運用する」としており、NII は、研究データ基盤の中核的なシステムとして、NII Research Data Cloud の整備を進めている。今後、これを着実に構築し、継続的に運用することが求められる。

また、研究データ基盤による研究データの利活用を進めるためには、研究データ基盤上で研究データを横断的かつ効率的に共有・公開・検索が可能となる仕組みを構築することが重要である。そのためには、研究データのメタデータを付与し、共通の識別子を登録して、研究データ基盤システム上で運用可能とすることが必要となる。

当面は 2020 年度の研究データ基盤システムの運用開始を目処として、図 2 に示す体制の下、各機関等から公開される研究データが、統一した識別子とメタデータのもとで横断的に検索可能となることを目指す。

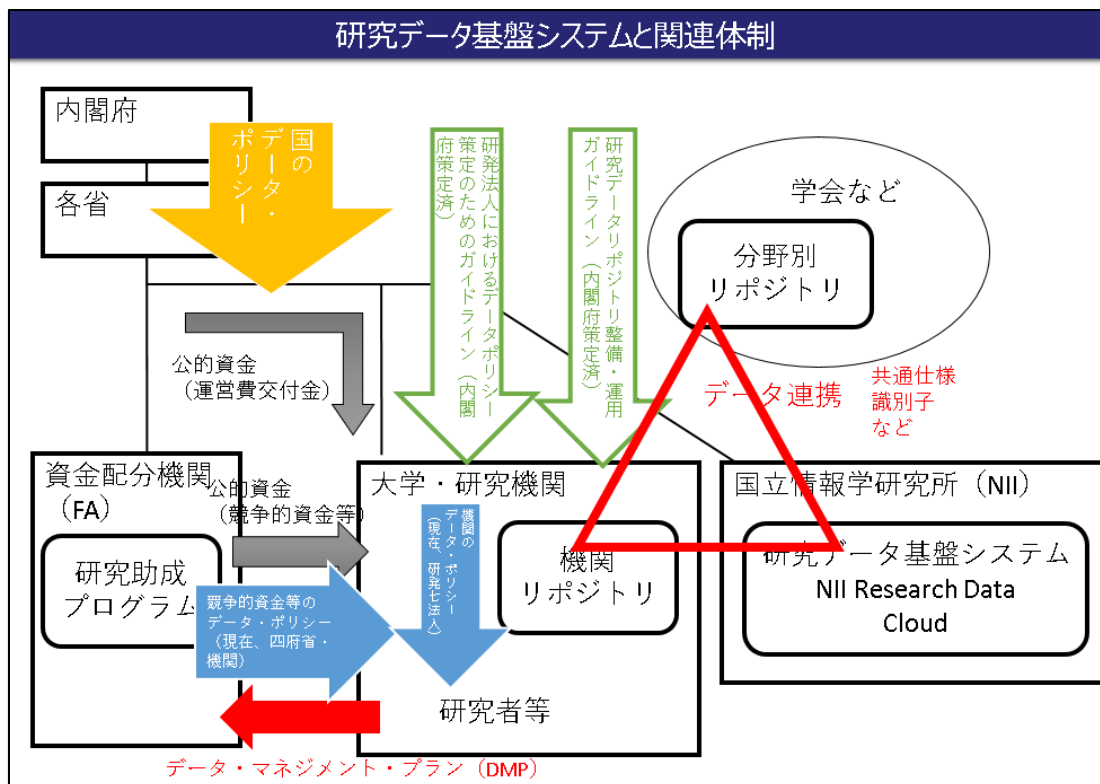


図 2 研究データ基盤システムと周辺環境の整備に関する体制

また、統合イノベーション戦略 2018 で示された、データポリシーや DMP の互換性の確保、共通語彙や識別子の統一等を通じて、Society 5.0 連携基盤やエビデンス・システムとの連携を進める。

3-2. 先進的な研究データマネジメントの推進

内閣府、文部科学省、経済産業省は、従来技術の延長にない、より大胆な発想に基づく挑戦的な研究開発を推進する「ムーンショット型研究開発制度」を創設した。

同制度では、未来社会を展望し、困難だが実現すれば大きなインパクトが期待される社会課題等を対象として、国が野心的な目標を掲げ、最先端研究をリードするトップ研究者等の指揮の下、世界中から研究者の英知を結集し、目標の実現を目指す。その際、我が国の基礎研究力を最大限に引き出す挑戦的な研究開発を積極的に推進し、失敗も許容しながら革新的な研究成果を発掘・育成に導く。また、同制度におけるマネジメントの方法については、進化する世界の研究開発動向を常に意識しながら、関係する研究開発全体を俯瞰して体制や内容を柔軟に見直すことができる形に刷新するとともに、最先端の研究支援システムを構築すること、また、将来の事業化を見据え、オープン・アンド・クローズ戦略の徹底を図ること等を旨としている。

以上の特徴に鑑み、公的資金による研究活動により生み出された研究データの利活用とそれによる先進的な研究マネジメント支援を促進するため、同制度において、先行的に研究データ基盤システムの活用を図るなど、先進的なデータマネジメントを推進する。研究データ基盤システムは、研究データの適切な管理や、研究プロジェクトの内部関係者や外部関係者とのデータ共有、また第三者を対象とした公開を柔軟に実施することなどを可能にするため、同制度の先進的な研究マネジメントを支援できるものと期待される。

また、同制度における先進的な研究データマネジメントや、国内外の優良事例の収集等により、研究データの類型化や、研究データ基盤の利活用を含む研究データの管理及び利活用のあり方の検討を進める。

3-3. 研究者へのインセンティブ

研究者が研究データ基盤システムを活用して、自発的に研究データの管理やメタデータの付与等を行い、また、積極的に研究データの利活用を図る等するためには、それらの活動を通じて研究者が便益を得られるというインセンティブが必要である。また、研究データ基盤システムの利用に際しては、適正な利用料により持続的な運営が可能な体制を構築することが必要である。

研究者へのインセンティブとしては、例えば、以下のようなものが考えられる。

- 研究開発プロセスにおける利便性の向上（実験・観測データや研究ノートの管理における研究者の負担軽減、プロジェクト関係者との柔軟なデータ共有設定など）
- 研究プロジェクト終了後も含めた、研究データの保存場所の提供

- 合理的な保存・管理経費
- 研究データの管理及び利活用に対する積極的な評価（新たなポストへの応募時において、論文リストと同様に研究データリストがこれまでの研究業績として評価される仕組みなど）
- 研究データの適切な経済的な価値付け
- 研究データの市場への流通
- セキュリティの確保
- 研究公正のエビデンスとしてのデータの蓄積
- 学術コミュニティからの支持

4. 人材の確保・育成

研究データの管理及び利活用に関しては、実験・観測データの体系的なマネジメントやメタデータの付与等の専門的な知見が必要である。これまでも、NII や各機関において、研究者等を対象とした研修活動も行ってきたところである。

今後、研究データの管理に関する業務が増大することを考慮すると、研究活動に対して知識や経験を有するシニア人材を活用することや、データ関係の支援業務を行う大学院生等に適切な報酬などのインセンティブを与えることなどが考えられる。また、既存の図書館業務の変革の中で、データ関係の支援業務を行う人材（データライブラリアン¹³、データキュレーター¹⁴、データアーキビスト¹⁵等）を充実していくことも望ましいと考えられる。

5. 国際展開

G7 茨城・つくば科学技術大臣会合においてオープンサイエンス WG が設置され、我が国は EU とともに共同議長を務めている。このことと関連して、本年 6 月には G7 による本 WG に関するワークショップが実施され、我が国としては、EOSC を擁する EU と積極的な協力関係の構築を図ることが確認された。また、2020 年度中に本格運用が予定されている研究基盤データシステム（NII Research Data Cloud）の開発に当たっては、欧州原子核研究機構（CERN）、米国 COS(Center for Open Science) 等と連携して進められてきた。これらを更に進め、データポリシーや技術、国際認証等の互換性を確保して、グローバルなデータプラットフォームの構築に向け、欧州のみならず、米、豪、加等との連携を深めていく。

また、以上のような国際的ネットワークの中で、我が国は、将来的には、相互運用性を確保したグローバルな研究データに関する情報基盤の構築に貢献し、我が国の研究データ基盤が日本ノードとしての地位を築くことで、世界的な商業出版社、IT・データ関係企業等との戦略的な連携関係を構築することを目指す。

¹³ データの整理のために、メタデータ付与やデータ連携などを安定的に行う人材。

¹⁴ データの利活用のために、コラボレーションなども利用して潜在的な価値を顕在化させる人材。

¹⁵ データの長期保存のために、データの管理・取得・廃棄などに関する計画を策定・実行する人材。

この他、研究データ基盤や研究データ利活用に関する国際的な枠組・組織（RDA¹⁶、WDS¹⁷、CODATA¹⁸等）において、我が国のプレゼンスを示し国際的な議論や合意醸成に参加していくことは重要であり、国としては、国内関係者のこうした活動を積極的に支援する。

6. 研究データ基盤整備と国際展開に関する目標

ここまでに記載した研究データ基盤整備ならびにその国際展開に関する活動については、以下の通り、2020年までの短期目標と、2025年までの中長期目標に分けて推進するものとする。

<短期目標：第5期科学技術基本計画期間（～2020年）>

- 研究データ基盤システム（NII Research Data Cloud）の運用開始（2020年度中）
- ムーンショット型研究開発制度における先進的な研究データマネジメントの実施（設計・運用の開始）
- メタデータの集約・横断的な検索体制の構築
 - データマネジメントに関する運用ガイドライン、データカタログなどの充実
 - 研究データ基盤システムの利活用の体制整備
 - 共通仕様、識別子等に基づくDMPなど
- 欧州（EOSC）等との連携

<中長期目標：第6期科学技術基本計画期間（～2025年）>

- 研究データ基盤システム（NII Research Data Cloud）の持続的体制構築
- ムーンショット型研究開発制度における先進的な研究データマネジメントの実施（運用・経験の蓄積・終了後の見直し）
- 共有、非公開データ等も含めたマネジメントの確立（範囲、手法などを含めて、国内外の優良事例の収集やムーンショット型研究開発制度における先進的な研究データマネジメントの実施を踏まえ検討）
- データ利活用情報の管理（研究データの利活用の研究開発への貢献度の測定等）とシステム・体制の評価及び改良
- グローバルプラットフォームへの拡張

7. 体制と役割

内閣府・CSTI：政府全体のデータ・ポリシー・ガイドラインの策定、フォローアップ
政府全体の国際窓口としての対応

日本学術会議：アカデミアの立場からの研究データ利活用のあり方の検討

¹⁶ Research Data Alliance（研究データ同盟）の略。

¹⁷ Word Data System（世界科学データシステム）の略。

¹⁸ Committee on Data for Science（科学技術データ委員会）の略。

文部科学省

NII：研究データ基盤システムの構築及び持続的運用、技術支援、研修の実施

NISTEP：研究者を対象とした実態調査

資金配分機関：研究助成プログラムのデータポリシーの策定、DMPの管理

大学・研究開発法人：機関のデータポリシーの策定、リポジトリの整備

その他 各府省（研究開発法人及び資金配分機関所管として）

資金配分機関：研究助成プログラムのデータポリシーの策定、DMPの管理

研究開発法人：機関のデータポリシーの策定、リポジトリの整備

8. 今後の課題

近年、通常の学術ジャーナル出版からアーカイブへとパワーシフトが生じている様に、データの重要性が最近強く意識されるに至り、今後も多様な変化が予見される。このような情勢を正確に判断しつつ、我が国の科学技術の発展とイノベーションの創出に向けたデータ基盤戦略の検討を継続的に実施することが肝要である。これまでの本 WG の検討を踏まえた今後の課題は以下のとおりである。

- ナショナル・データ・ポリシーの策定（内閣府が主導し本年中目処で策定）
- 研究データの帰属（オーナーシップ）や管理主体に関する法的な位置づけ
- 研究者が研究データの保存及び管理を積極的に行うための具体的なインセンティブ
- 研究データの経済的な価値付けとデータ利活用に関するビジネスモデルなど（研究データ基盤と関連する体制の持続と民間企業との連携など）
- 産学官におけるデータ連携のための仕組み・環境の整備（相互運用性の向上）
- 研究データ関連支援業務を担う人材の確保・育成に関する具体的な方策
- 研究データ管理及び利活用に関する内外の成功事例の収集、ムーンショット研究開発制度における先進的な研究データマネジメント等による経験の蓄積
- 国際展開の方策（グローバルな研究データに関するプラットフォームの拡張に向けた協力、出版社・IT・データ関係企業等との戦略的提携、国間の相互利益（レシプロシティ）の確保など）

など。

(別添) 国レベルのデータ・ポリシー（ガイドライン）で規定すべき項目

○研究データのマネジメントに関する基本的考え方

- ・ 公的資金による研究データの管理及び利活用に関する基本的な考え方の提示
- ・ 研究データを公開・共有・非公開とする場合の要件
- ・ 研究データの区分と求められる要件
- ・ 経費の性格毎のデータ管理のあり方
- ・ 機関や研究費の特性による研究データの管理のあり方

など

○研究データ基盤の構築と体制整備

- ・ 求められる研究データ基盤の要件
- ・ 研究データの横断的検索が可能となるための技術的・制度的事項（識別子、メタデータの仕様の共通化など）

など

○公的資金に関する各ステークホルダーの責務

- ・ 内閣府及び各省の役割
- ・ 責任機関の役割
- ・ 資金配分機関の役割
- ・ アカデミアの役割
- ・ 研究者の役割

○その他

（参考 1）研究データの範囲

○内閣府「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～」(平成 31 年 4 月 2 日)

Q 5. データポリシーにおいて「機関として管理の対象とするデータ」の定義はどのように整理したらよいか。

A 5. 研究データの定義は各機関によって異なるが、概ね「研究の過程、あるいは研究の結果として収集・生成される情報」と定義でき、観測データや実験データ、シミュレーションを行った結果得られたデータなどを含む。公開・共有するデータに加え、どのようなデータを機関として管理するのかについては、各研究分野のデータ管理の状況も踏まえ、各法人での判断が必要となる。データポリシー策定済み法人の定義・範囲なども参考に各法人でご検討いただきたい。

○経済産業省「委託研究開発におけるデータマネジメントに関する運用ガイドライン」(平成 29 年 12 月) <抜粋>

<適用対象データ>

(1) 適用対象事業における研究開発で取得又は収集した電磁的記録(電子的方式、磁気的方式その他人の知覚によっては認識することができない方式で作られる記録であって、電子計算機による情報処理の用に供されるものをいう。)を対象とする。

(2) 適用対象データとしては、例えば、開発する技術や製品の性能等を示すための計測データ、センサ等から取得した計測・収集データ、材料物性データ、製造条件データ、材料構造データ、機械学習やシミュレーション・プログラムにインプットするデータ、学習用データ(データの集合体)、学習済みモデルなどの様々なデータを対象とする。

(2) 研究開発データの分類化

研究開発データはその性質により、その取得又は収集目的や利活用すべき者が異なる。したがって、研究開発データを、委託者が管理すべき研究開発データであり、委託者に提供される研究開発データとして指定された研究開発データ(以下「委託者指定データ」という。)、委託者指定データ以外の研究開発データであって、プロジェクト参加者が自主的に管理する研究開発データ(以下「自主管理データ」という。)、委託者指定データ又は自主管理データ以外の特に管理を要しない研究開発データ(以下「非管理データ」という。)の3つの研究開発データに分類してそれぞれのマネジメントを検討する必要がある。

○AXIES 大学 ICT 推進協議会「学術機関における研究データ管理に関する提言」

研究データ (research data)

研究活動の過程で生み出されるデータ。研究素材としての生データだけでなく、その中間過程の加工データや、研究成果発表に通ずる解析データ、研究成果発表の根拠データも含み、またそれらデータを説明する資料も含む。データ取得条件やソースコードなど、研究データの生成を再現可能とするための研究環境に関わる情報を含む場合もある。併せて、学会発表

や学术论文、研究プロジェクト評価書などを含む研究成果物、当該研究に関連する研究助成や各種調達に関わる文書、共同研究者等との間のコミュニケーションに関わる記録、広報関連資料など、学術研究活動に付随して生成・作成されるあらゆる資料を含むものの総称として用いる場合もある。研究データの形態としては、数値、画像、テキストなど、あらゆる形態をとりうる。なお、本提言はデジタル時代における研究データ管理の可能性の進展を想定しているため、これらのなかでも「デジタルデータ」を中心的に想定する。物理的な形をとる研究資料や文書等については、これらを説明するデジタルデータを媒介として、対象に組み入れることができる。研究データにどこまでを含めるかについては、各機関が定める研究データ管理の導入目的とスコープに応じて、各機関が定める必要がある。

(参考2) 研究データの管理・利活用に関連する主な法令及びガイドライン

- ① 輸出貿易管理令（昭和二十四年政令第三百七十八号）

- ② 補助金等に係る予算の執行の適正化に関する法律（昭和三十年法律第七十九号）

- ③ 不正競争防止法（平成五年法律第四十七号）

- ④ 不正アクセス行為の禁止等に関する法律（平成十一年法律第二百二十八号）

- ⑤ 個人情報の保護に関する法律（平成十五年法律第五十七号）

- ⑥ 産業競争力強化法（バイドール）（平成二十五年法律第九十八号）

- ⑦ 委託研究開発におけるデータマネジメントに関する運用ガイドライン
担当部局：経済産業省 産業技術環境局
公表日：平成29年12月27日（水）

- ⑧ AI・データの利用に関する契約ガイドライン—データ編—
担当部局：経済産業省 商務情報政策局
公表日：平成30年6月15日（金）

(参考3) DMPの一般的な項目

項目	備考
1. プロジェクト情報	
・事業年度	
・事業名	
・研究開発課題名 等	
2. データ情報	
・取得・収集される研究開発データの名称	データ種別、フォーマット、容量 等
3. 担当者情報	
・研究代表者	名称、識別子 (NRID、ORCID 等)、所属機関 等
・研究分担者	同上
・研究開発データの管理者	同上
・研究開発データの取得者又は収集者	同上
・(仮) データキュレーター (データ専門職)	同上、データキュレーターの定義が必要
4. 取得・収集	
・研究開発データの取得又は収集方法	関連する標準や方法、品質保証、データの組織化 (命名規則、バージョン管理、フォルダ構造) 等
5. 文書化とメタデータ	
・再利用を可能にするための文書	データの収集・分析・処理方法に関する情報、変数の定義、測定単位、設定した仮説、データのフォーマットとファイル種別、使用したソフトウェア 等
・使用するメタデータ標準 (フォーマット)	汎用 (Dublin Core、DataCite、JPCOAR スキーマ 等) 分野 (DDI (社会科学)、SPASE (観測データ) 等)
6. 研究プロセス中の保管とバックアップ	
・研究中のデータおよびメタデータの保管とバックアップ方法	保管とバックアップの場所や、バックアップの頻度
・研究中のデータのセキュリティと機密データの保護方法	インシデント発生時のデータ復旧方法 アクセス可能なメンバー、アクセスの管理方法 機密データのリスクや管理方法 どの機関のデータ保護方針が適用されるか

7. 倫理・法律のコンプライアンス	
・倫理的問題への対処	データの保存や共有に関する同意、匿名化処理、センシティブデータの扱い 等
・著作権や知的財産権に関する問題への対処	データの所有者は誰か、再利用のためのライセンス付与、第三者による再利用の制約 等
8. 研究プロセス後の保存・管理	
・研究開発データの保存・管理方針	保存すべきデータの選定 等
・保存期間	日本学術会議は研究公正の観点から 10 年程度
・保存場所	リポジトリ、ストレージ
9. 公開・提供	
・研究開発データの公開・提供方針	公開・限定的公開・非公開の方針 非公開の場合の期間、理由 等
・研究開発データの公開・提供場所	リポジトリ（汎用、分野、機関）
・委託者指定データ、自主管理データの分類	NEDO では、交通情報データ、人口衛星画像データなどを委託者指定データとして例示

(参考4) 各資金配分機関のDMPの比較

大項目	詳細	AMED	JST	NEDO
プロジェクト 情報	事業年度	○	○	○
	事業名	○	○	○
	研究開発課題名	○	○	○
データ情報	研究から産出されるデータ及びデータ群の総称	○	○	○
	委託者指定データ、自主管理データの分類			○
	研究データの種類の分類		○	
	研究開発データの説明		○	○
	想定データ量			○
	加工方針（ファイル形式、メタデータに関する事項を含む）			○
	研究計画書対応項目		○	
担当者	データサイエンティストの所属・氏名等	○		
	研究開発データを取得又は収集した者			○
	研究開発データの管理者			○
取得・収集	研究開発データの取得又は収集方法			○
保存・管理	リポジトリ（保存場所）	○	○	○
	管理対象となる研究データの保存・管理方針		○	○
公開・提供	研究データの公開・非公開に係る方針		○	○
	公開可能な研究データの提供方法・体制		○	○
	公開研究データの想定利用用途		○	○
	公開研究データの利活用促進に向けた取り組み		○	○
	研究開発データの利活用・提供方針			○
	（他者に提供する場合）円滑な提供に向けた取り組み； （秘匿して自ら利活用する場合）秘匿期間、秘匿理由			○

※なお、表中のAMEDのDMPは汎用的なDMPについてであり、ゲノム関連プロジェクトにおいては、データシェアリングポリシーを設け詳細なDMPを義務づけている。

(参考5) DCC (Data Curation Centre ; 英) の DMP

大項目	詳細	備考
管理上のデータ	助成機関や所属機関が定めた関連 ID	
	助成機関・助成金番号	
	プロジェクト名称・内容・連絡先	
	プロジェクトの主任研究員 (PI) もしくは主要研究者の名前とその永続的識別子 (ORCID 等)	
	データ管理計画書の作成日・最終更新日	
	関連するポリシー	
データ収集	どのようなデータを収集・生成するのか	データの種別、フォーマット、容量 etc.
		フォーマットやソフトウェアのデータ共有・長期アクセスへの適性
	どのようにデータを収集・生成するのか	再利用できる既存のデータの有無
		用いる標準や方法論
文書化とメタデータの付与	文書化とメタデータの付与	フォルダやファイルの構造化の方法
		バージョン管理の方法
		長期にわたりデータが可読性を有し解釈され得るために必要な情報
倫理・法律のコンプライアンス	倫理的問題	文書化及びメタデータ作成の方法
		用いるメタデータ標準とその理由
		データの保存、共有についての同意取得の有無
	著作権、知的財産権等に関する問題	(必要な場合の) 参加者の個人識別性の保護
		データの所有者
		センシティブデータの安全な保管・移動のためのとりくみ
保管とバックアップ	研究期間中のデータの保管・バックアップ方法	再利用のために付与するライセンス
		データの所有者
		第三者による再利用についての制限の有無
保管とバックアップ	研究期間中のデータの保管・バックアップ方法	十分なストレージを有しているか
		付加サービスの利用に必要な料金を使途に含める必要があるか
		データのバックアップ方法

		障害発生時のデータの復元方法
	アクセス制限やセキュリティ確保の管理	
選定と保存	どのデータを保持・共有・保存すべきか	契約、法律、規制等の理由で保持あるいは破棄すべきデータ
		その他の保存するデータの決定方法
		データの保持・保存の期間
	データセットの長期保存計画の内容	リポジトリやアーカイブ等のデータの保存場所
データ共有	どのように共有するか	
	データ共有に何らかの制限が必要か	
責任とリソース	データ管理の責任者	
	計画実行のために、どのようなリソースが必要か	

(参考6) 各研究機関におけるデータマネジメントの実例

		JAMSTEC		NIMS	AIST		理研
事業内容		津軽海峡東部海洋レーダーデータサイト	船舶観測データ・サンプルの計測・取得	材料データプラットフォーム	人体形状の計測	歩行データの計測	FANTOMプロジェクト*
データ公開状況	生データ	一部公開：レーダーデータは利用申請をすることでダウンロード可能。水温データは Web サイトでグラフを公開、ダウンロード不可。	公開：航海・潜航に関する品質管理を施したデータ	データごとに適切な公開条件を付して公開 (利活用を図るために登録されるデータ)	公開：人間計測データ	公開：歩行データ 非公開：転倒歴など	公開：独自技術 (Cap-Trap 法、CAGE 法) より得られた完全長 cDNA 配列や RNA の 5' 配列データ
	加工データ	公開：レーダーデータから算出した海表面の流向・流速データ	—	データごとに適切な公開条件を付して公開 (研究成果物として公表されたもの)	非公開：解析・二次データ 有償公開：統計量	非公開：転倒スコアなど	公開：哺乳類電子の機能情報 (遺伝子の位置や発現機構など) データファイルならびに専用ビューワーを介して提供
	メタデータ	公開	公開	公開	公開	公開	公開

メタデータの 内容	<ul style="list-style-type: none"> データ取得日 データ取得方法 データ取得者 	<ul style="list-style-type: none"> データ取得日 データ位置情報 データ取得方法 データ処理方法 データ処理者 データ形式 キーワード 	<ul style="list-style-type: none"> データ作成日、公開日等 公開と利用の条件 データ作者情報 データ来歴、装置、試料等の情報 	<ul style="list-style-type: none"> データ取得日 データ取得方法・項目 実験対象者の性別・年齢 データ形式 	<ul style="list-style-type: none"> データ取得期間 データ取得方法・項目 実験対象者の性別・年齢・身長・体重 データ形式 	<ul style="list-style-type: none"> 使用サンプルの詳細情報 サンプルの分類情報 配列データの取得方法 データ形式
データ公開 猶予期間	30分ごとの準リアルタイム	<ul style="list-style-type: none"> 航海メタデータは原則2ヶ月 データは原則2年。ただし、定常観測データは処理が済み次第公開 	データポリシーおよびデータリポジトリ運用規則に沿ってデータ登録者が設定する期間	既に公開済み	既に公開済み	論文発表と同時公開 (既に公開済み)
データ公開 猶予理由	—	学会発表、論文作成、特許作成	学会発表、論文作成、特許作成など			論文作成のため
第三者機関による データ利活用の有償・無償基準	<ul style="list-style-type: none"> 科学的・教育的目的利用は無償 産業利用は有償 	<ul style="list-style-type: none"> 科学的・教育的目的利用は無償 産業利用は有償 	原則無償(データポリシーおよびデータリポジトリ運用規則に沿って有償等の可能性あり)	無償	無償	無償

公開媒体	Web サイト	<ul style="list-style-type: none"> ・Web サイト(メタデータ・データファイル) ・オフライン提供(生データ、高解像度映像、サンプル) 	データリポジトリ	ホームページ	ホームページ	<ul style="list-style-type: none"> ・専用 Web サイト(理研内) ・国際塩基配列 DB コレクション ・公開リポジトリサービス ・NBDC (JST 内) の生命科学系データベースアーカイブ
データ帰属先	<ul style="list-style-type: none"> ・レーダーデータは JAMSTEC に帰属 ・水温データの一部は JAMSTEC に帰属、一部は観測を実施している他機関に帰属 	原則 JAMSTEC に帰属	NIMS に帰属。ただし共同研究契約等がある場合はその契約に依る	原則 AIST に帰属	原則 AIST に帰属	理研に帰属するが、データは CC BY 4.0 で公開

*) 理研主導による国際共同研究プロジェクト (FANTOM = Functional Annotation Of Mammalian genome)。哺乳類を対象として、網羅的に遺伝子の機能 (ゲノム中の遺伝子の位置や発現機構) を解析。