

国立研究開発法人理化学研究所 開拓研究本部 研究データの管理、公開、共有及び利用に関する実施ガイドライン（開拓研究本部研究データガイドライン）

令和3年3月22日

開拓研究本部

1. 目的

国立研究開発法人理化学研究所（以下、研究所）開拓研究本部は、抜きんできた基礎研究成果を生み出すとともに、理研内外の研究者を有機的に連携する分野横断的な研究プロジェクトを推進することにより、新たな科学の創成を進め、日本における戦略的研究プロジェクトの芽となる研究を開拓する組織である。

開拓研究本部は、研究所が定める「研究データの管理、公開、共有及び利用に関する基本方針」（以下、基本方針）を継承し研究成果最大化に貢献する。開拓研究本部研究データガイドライン（以下、本ガイドライン）は、開拓研究本部に所属する研究者等が適切かつ確実に研究データの管理が行えるよう、開拓研究本部における研究データの取り扱いに必要な事項を定義することを目的とする。

2. 研究データおよび利活用データの定義

[研究データ] 基本方針において、研究データとは研究所の研究活動を通じて取得、作成された 1.目的に合致するデータをいう。開拓研究本部においては、原則研究データを電磁的な方法により管理する。研究データ以外の研究記録、研究成果有体物等は、研究記録管理規程、研究成果有体物取扱規程等の規定により取扱が規定されているものは、それら規定に従うものとする。研究データには、その利用、収集する状況、由来等に応じて、以下に掲げる呼称が与えられる。

[一次データ] 研究の過程で新規に産出される研究データを一次データという。一次データには、観測機器、計測機器、撮像装置、計算機等の研究機器から生成され収集されたメタデータを含むデータ、研究者等自身が計測、考案して作成した研究データ等が含まれる。

[二次データ] 一次データ、あるいは二次データ、あるいはその双方を加工して作成されたデータを二次データという。

[利活用データ] 基本方針の定義に従い、メタデータを含む研究データのうち、研究記録管理規程第 5 条で保管が義務付けられたデータ、利活用によって科学技術の発展に貢献できると研究所を代表して開拓研究本部のデータ管理者が判断したデータを「利活用データ」という。

[ローカル研究データ] 研究者等あるいは研究プロジェクトが専ら使用し、それ以外の者からアクセスできないデータ管理領域に置かれる研究データをローカル研究データという。

[ローカル利活用データ] ローカル研究データのうち利活用データであるものをローカル利活用データという。

[データ生成者] 研究データを生成した者をいう。

[データ管理者] データ生成者により生成された研究データを、情報基盤や情報媒体を用いて管理する者で、所属長が指名する者をいう。データ管理者は上記データ生成者と同一の者であってもよいし、別の者であってもよい。

[データ利用者] 利活用データを受け取り利用する者をいう。

3. 研究データリポジトリの選定及び利用

利活用データのみならず研究データの管理には、原則研究所が整備する研究データリポジトリ（以下、理研リポジトリ）を利用する。研究データを他の管理領域に保管する場合は、そのデータへのリンクを理研リポジトリに登録することが望ましい。

理研リポジトリ機能の増強を研究所に要請する場合、それには客観性、定量性を持った根拠が必要となる。開拓研究本部は常に理研リポジトリの実効的な運用が行えるよう、開拓研究本部内で産出される研究データの質や量を正確に把握できるよう努める。

4. 人を対象とする医学系研究における個人情報等の取り扱い

人を対象とする医学系研究の実施に当たり、開拓研究本部内で生成された、あるいは

研究所外から受領した個人情報を含む研究データについては、機微性が極めて高い要配慮情報として厳重に取り扱う。関係法令・ガイドライン・倫理指針等を遵守するほか、研究所が定める個人情報等の取り扱い規程等に従う。さらに研究所外から受領したデータについては、上記規定等を含む情報管理に関するガイドラインに沿って処理を施したのち、研究に供する一次データとして扱う。

5. 利活用データの選択と管理

研究データから利活用データを選択する際に、データ生成者あるいはデータ管理者が、データ保管あるいは他者とのデータ共有の必要性を、研究の状況、研究知財の管理、論文根拠としての必要性等を含む多様な側面から、研究者としての良心に従って総合的に評価して判断する。具体的には、利活用データとして選択されるデータとして、

- ・ラボノートに記載があり、保管が必要となっているデータ
- ・共同研究で共有が予定されているデータ
- ・研究の次の段階で必要となるデータ
- ・将来執筆予定の論文掲載、根拠に必要なデータ

などが挙げられる。

基本方針が定める「リポジトリに登録される利活用データの保管期間」は、登録されてから原則 10 年間以上である（ローカル研究データとしての保管期間を含み、合算するものとする）。

ここで、研究データの保管には理研リポジトリの管理、維持のためには費用がかかることも考慮する。利活用データを、データ流通や利活用に支障のない範囲で割愛、データ圧縮、プログラムで生成可能なデータはその元データのみに限る等の検討を積極的に行い、費用削減による実効性向上に努める。

ローカル研究データ、ローカル利活用データについては、当該のデータ管理者が責任をもって管理する。ローカル利活用データの保存期間を 10 年とする。それらデータを理研リポジトリ以外の情報基盤や媒体で管理する場合には、情報媒体の盗難、紛失、第三者からのアクセスを防止する措置、廃棄時のデータ漏洩対策を適切に施す。

6. 公開共有フラグの変更

「公開共有フラグ」とは、理研リポジトリに登録された利活用データについて、その公開先を指定するための情報を公開共有フラグという。基本方針に従い、公開共有フラグには、「非公開」、「所内限定公開」、「所内外限定公開」、「公開」が含まれる。

公開共有フラグ「非公開」を付して理研リポジトリに登録された利活用データは、データ生成者あるいはデータ管理者が、研究動向やデータ利用者を見極め、公開共有フラグの変更を所属長の承認を得て行う。公開共有フラグは、対象となるデータ利用者の範囲が狭い順に列挙すると以下の通りとなる。

[非公開] 研究者等あるいは研究プロジェクトのみがアクセス可能なデータに付与される。

[所内限定公開] 研究所内のデータ利用者が読み込むことができるデータに付与される。研究所内の者すべてを対象とすることも、パスワードや IP アドレスによるアクセス制限などの手法を適用してその中の一部の者のみを対象とすることもできる。

[所内外限定公開] 研究所内外を問わず、単数あるいは複数のデータ利用者を対象にしたデータに付与される。

[公開] データ利用者を限定せず、何人も読み込み可能なデータに付与される。

ここで、データ利用者へのデータ提供の手段については、データ特性等に応じて個別に選択する。例えば、常にネットワーク等を通じて読み込み可能な状態にしておく必要はなく、書面を通じて読み込み依頼を受け、情報媒体を用いて提供することも可能とする。

公開共有フラグは設定後、常に研究動向、利活用データの価値、共同研究の進展、公開の有効性等を総合的に勘案し、適切な設定となるよう努める。

7. 利活用データの公開猶予期間

利活用データは、当該データが生成された時点から 2 年以内あるいは利活用データとして選択された時点から 1 年以内のいずれか早い時点で、原則公開する。しかしなが

ら、当該利活用データに関係する契約や論文掲載条件等の取り決めがある場合には個別に対応する。

8. 利活用データの利用許諾条件

公開共有の範囲すなわち公開共有フラグと連動し、データ著作権者の権利が侵害されないよう適切な利用許諾条件をデータ利用者に示すよう努める。共同研究契約やMTA(Material Transfer Agreement)が締結されている場合には、それに従った利用許諾を求める。データ利用者を特定しない場合でデータの利活用状況を把握したい場合には、データをダウンロードする前に IP アドレス等の記録を求める。もしくは受取先から個人情報であることに留意しながらデータ利用者の氏名や所属などのデータ利用者情報の提供を求める。さらには、データ利用者でデータを二次利用の成果物として論文等で発表する場合には、論文の本文や謝辞にて提供元情報の掲載を義務付ける利用許諾条件を明記する。利用許諾条件の作成に当たっては、データ生成者やデータを参照する識別子を論文に掲載するように求める条文を含める等、論文の読者が引用元データを参照できるよう配慮する。一方で、プログラムからのデータ利用なども含め、障壁なくデータの読み込みを可能としてより多くの受取先にデータを提供したい場合には、オープンデータに適した利用許諾条件、例えばクリエイティブコモンズライセンスを利用する。

9. 公共リポジトリの利用

公開データの取り扱いにおいては、研究分野やデータ種によっては公共リポジトリの利用が好ましい場合がある。公共リポジトリが定めるデータ公開期間や利用許諾等の公開条件を総合的に勘案し、公共リポジトリの利用が適切な場合には公共リポジトリに公開データを登録する。この時、メタデータについては研究所が整備する理研リポジトリに登録し公開する。

10. メタデータの付与

既に大小を問わず国内外の研究機関がウェブ上でメタデータの公開を開始しており、メタデータがつなぐ研究データのグローバルネットワークが形成されつつある。開拓

研究本部もこの潮流に準じて、オープンサイエンスや研究データ利活用推進のため、利用価値の高い研究データを選んで高品質なメタデータを付与するよう努める。メタデータの記述には、グローバルな相互運用性を担保する点でも標準化されたオントロジーやデータ項目を活用することが望まれる。データ生成者は、情報統合本部の担当部署が提供、推奨するオントロジーやツール類の活用、技術支援を得ながら世界最高品質のメタデータ作成に努める。

メタデータの生成あるいは作成は、対象となるデータが生成された時点で行うことが、データ管理のコストや便宜を考慮すれば最も好適である。しなしながら、利活用データの選定時やデータ共有・公開時に付与されることも多く、メタデータの生成に要する労力がより多くなる場合が多い。このため、研究計画の段階でメタデータの詳細度、付与する時期等も考慮しながら研究データの取り扱い計画を立案するよう努める。

メタデータの記述が実験条件、観察された現象等の科学的知見の詳細に及ぶ場合、メタデータ自体が価値の高い研究データとなりうる。この場合には、メタデータではない利活用データと同様に、適切な公開共有フラグや利用許諾条件が付されるよう努める。しかしながらデータのカatalogとして作成されたメタデータは、プログラム等から機械的にメタデータ検索が行えるようなクエリの標準技術の普及を考慮して、所属長の承認の下でクリエイティブコモンズライセンスを付し、アクセス障壁を低く抑えながらデータ公開を行う。

データとそれを説明するメタデータとは密接な関係にあるが、公開共有の設定は異なってもよい場合も考慮する。例えば、データの知財を保護しながらもメタデータを活用して研究内容の公開を積極的に行うことを考慮して、メタデータは公開であるが、データは非公開であることも許容される。さらに、データ公開を緊急に行う場合で、詳細なメタデータの付与に時間を要する場合には、データ生成者あるいはデータの所在等の最低限のメタデータを付与してデータを先行的に公開することも許容される。

11. データ管理者の転出等対応

理研リポジトリに登録された利活用データの管理は、データ管理者あるいはデータ管理者の所属長が責任を持って行う。データ管理者が開拓研究本部から転出する場合には、データ管理者の所属長が管理責任を請け負う。ここで請け負った所属長は、別の

担当者に管理を依頼することができる。誰も管理責任を負えない状況となった利活用データについては、開拓研究推進室に相談し、その管理方法について開拓研究本部長が決定する。

12. データの削除

基本方針に従い、メタデータを含む利活用データは10年間を保管期間とし、保管期間終了後は削除する。しかしながら、学術上の重要性を鑑みて保管期間を延長することが望まれる場合には、所属長の承認を得て、期限をつけて保管期間を延長する。

他方、保管期間前に削除することが妥当と判断される利活用データは、妥当性を結論づける理由を添えて所属長の承認を得たうえで削除するとともに、その理由を理研リポジトリ上に削除前と同一の公開共有フラグを設定して登録する。

データの削除においては、データとそのメタデータとは必ずしも一体のものとして同時に削除する必要はなく、それぞれ適切な削除計画を立てるよう努める。

13. 研究者等の評価及び報奨

研究データを生成、取得し、さらに利活用データとしてデータ利用者と共有を積極的に行った者については、当該活動を業績として考慮し、適切な評価及び報奨を行う。

14. 施行日

本ガイドラインは、令和3年4月1日から施行する。施行前に生成していた研究データについては、速やかに「5. 利活用データの選択と管理」を基にし、利活用データの選択を行う。

以上