

# 国立研究開発法人理化学研究所 計算科学研究センター 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（計算科学研究センター研究データガイドライン）

（令和3年2月15日 計算科学研究センター会議決定）

## 1. 目的

国立研究開発法人理化学研究所（以下、単に「研究所」という。）計算科学研究センターは、スーパーコンピュータを中心とした高性能な「計算」という事象自身を「計算の科学」として探求し、それによって得られる莫大な計算パワーを様々な科学分野の問題解決に適用してそれらの発展に寄与する「計算による科学」を推進し、更には両者の高度化に貢献する他の科学分野の産物である「計算のための科学」と連携を果たすべく、次世代の「計算科学」の世界トップレベルかつ我が国の中核拠点の研究センターとして活動する組織である。

この計算科学研究センター研究データガイドライン（以下、「本ガイドライン」という。）は、計算科学研究センター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、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指針」（以下、基本指針）に加えて、計算科学研究センターにおける研究データの取り扱いに必要な事項を定めることを目的とする。

## 2. 研究データおよび利活用データに関連する用語の定義

[研究データ] 基本指針において定義する通り、研究データとは研究所の研究活動を通じて取得、作成されたあらゆるデータをいう。なお、計算科学研究センターにおいては、すべての研究データを電磁的な方法により管理する（紙媒体上のデータのような直接電磁的な方法で管理できないデータは電磁的な方法により管理できるよう適切に変換する）こととし、本ガイドラインにおいてはデジタル形式のもののみを対象とする。ただし、研究データには研究ノート、論文、およびデータベースシステムやソフトウェアコンポーネント等の計算機上で実行されるプログラム類は含めない。研究データには、その利用、収集する状況、由来等に応じて、以下に掲げる呼称を与える。

[一次データ] 研究の過程で新規に生成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、研究者自身が計測、考案して生成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生成されたデータを二次データという。

[利活用データ] 基本指針の定義に従い、メタデータを含む研究データのうち、論文発表等に伴い公開が義務付けられたデータ（以下、「論文データ」という。）、利活用によって科学技術の発展に貢献できると所属長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] 研究所の研究者あるいはその研究者の参画する研究プロジェクトのメンバーが専ら使用し、第三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。ローカル研究データの管理媒体には、研究者が専ら個人で利用する計算機に接続された固定記憶装置、種々のディスクや不揮発性メモリを備えた記憶メディア、ネットワーク経由でのアクセスが可能で専ら利用者のみがアクセスできる記憶領域等が挙げられる。

[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。

[公開共有フラグ] 研究データリポジトリに登録された利活用データについて、そのデータ利用者の範囲を指定するための情報を公開共有フラグという。基本指針に従い、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。

[データ生成者] 研究データを生成した者をいう。なお、二次データを生成した者が一次データを生成した者と異なる場合、権利の帰属の扱いについて留意が必要である。

[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」という。）を利用する。

しかしながら、大規模なデータが生成された場合等においては、理研リポジトリの容量や転送ネットワークの性能などの状況によっては実効的なデータ管理が行えない場合がある。この場合、理研リポジトリの機能の増強を研究所に要請することになるが、それには客観性、定量性を持った根拠が必要となる。計算科学研究センターは常に理研リポジトリの実効的な利用が行えるよう、計算科学研究センターで生成される研究データの質や量を正確に計測できるよう努める。

#### 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、計算科学研究センター内で生成された、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の取り扱いについては、「人を対象とする医学系研究における個人情報等の管理に関するガイドライン」に従う。

#### 5. 研究データリポジトリへの登録

研究の過程で生成、管理される研究データは、一次データの生成に始まり、共有や保管のために理研リポジトリへの登録、データ公開、データ削除に至る一連の流れが典型的なライフサイクルとなる。このライフサイクルをより具体的に記述すると以下の通りとなる。ここで、データ生成者とデータ管理者は異なっても同一の人物であってもよい。

##### a. 一次データの生成

データ生成者により生成された一次データを、データ管理者がローカル研究データとして管理する。

##### b. 二次データの生成

データ生成者が一次データあるいは二次データに解析・加工等の処理を施し、二次データを生成する。データ管理者は、生成された二次データはローカル研究データとして管理する。

##### c. 利活用データの選択

データ管理者がローカル研究データとして管理されている一次データあるいは二次デ

ータから、保管・共有・公開の対象となるデータを利活用データとして選択し、ローカル利活用データを生成する。

d. 理研リポジトリへの登録

データ管理者は、理研リポジトリが提供するアップロード機能を用いて、ローカル利活用データを理研リポジトリにアップロードし登録を行う。登録の際、不用意にデータ公開が行われないよう、公開共有フラグは「非公開」を設定しておくことが望ましい。

e. 公開共有フラグの設定及び変更

データ管理者は、後述 6.の手続きにより公開共有フラグを設定、変更する

f. データの削除

データ管理者は、後述 13.の手続きによりデータを削除する

研究データから利活用データを選択する際に、データ管理者は、データ生成者と相談しつつ、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。

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、それらのデータ管理者が責

任をもって管理する。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. 公開共有フラグの設定及び変更

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、公開共有フラグの設定及び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] データ生成者、データ管理者及びその所属長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、データ提供の手段について、データ管理者は、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、必要に応じた設定変更を所属長の承認の下でデータ管理者が行う。

## 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する研究実施時に契約や取り決めがある場合には個別に対応する。さらに論文データについては、論文掲載条件に論文データ公開猶予期間が規定されている場合には、当該規程に従う。

## 8. 利活用データの利用許諾

データ管理者は、6.のデータ利用者の範囲すなわち公開共有フラグの設定又は変更と連動し、研究データを第三者が取得・再利用する際に、無断転載や剽窃、不適切な加工、利害関係者とのトラブルといった問題の発生を防ぎ、データ生成者及びデータ管理者が望む利用条件が簡潔に伝わるよう、受取先に適切な利用許諾を示すよう努める。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。受取先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 アドレス等の記録とともに受取先から個人情報であることに留意しながら受取先の氏名や所属などの受取先情報の提供を求める。さらには、受取先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を明記する。利用許諾の作成に当たっては、データ作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、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. 利活用データの保管期間と理研リポジトリでの管理方法

基本指針に従い、理研リポジトリに登録されるメタデータを含む利活用データの保管期間は、登録されてから原則 10 年間とする。利活用データ保管期間中に科学技術が進歩あるいは潮流が変化する等の事情により、当該利活用データの利用価値は変動することが想定される。このため、データ管理者は、定期的に理研リポジトリに登録された利活用データを確認し、当該データに適合した公開共有フラグと利用許諾の設定に

努める。

## 10. 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。

## 11. 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。計算科学研究センター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。研究データを生成する研究者は、計算科学研究センターの担当部署が提供、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。

メタデータの生成は、対象となるデータが産出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。

## 12. データ管理者の転出対応

リポジトリに登録された利活用データの管理は、原則データ管理者が責任を持って行う。データ管理者が計算科学研究センター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る利活用データについては、計算科学研究センター副センター長（研究担当）が指定した部署に相談し、その管理方法について決定する。

## 13. データの削除

データ管理者は、所属長の承認の下、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許可を得たうえで理研リポジトリから削除するとともに、その理由を当該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。

## 14. 研究者等の評価及び報奨

研究データを生成し、さらに利活用データとして他者と共有を積極的に行った研究者等については、所属長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。

以上