

国立研究開発法人理化学研究所 バイオリソース研究センター 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 (BRC 研究データガイドライン)

令和 3 年 3 月 22 日

バイオリソース研究センター

目次

1. 目的	3
2. 本ガイドラインの改訂方法について.....	3
3. 本ガイドラインにおける用語の説明、定義.....	3
3.1 本ガイドラインにおける「データ」の定義.....	3
3.1.1 データ内容に基づく分類 (区別)	4
3.1.2 データのライフサイクル視点からの分類.....	6
3.1.3 その他のデータの分類.....	8
3.2 その他関連する用語.....	9
3.2.1 [公開共有の種類].....	9
3.2.2 [データ生成者].....	10
3.2.3 [データ管理者].....	10
3.2.4 [データ利用者].....	10
3.2.5 [データリポジトリ].....	10
3.2.6 [(情報の) トレーサビリティ].....	10
4. BRC における研究データの取り扱い、管理の指針	11
4.1 研究データ取り扱いの指針	11
4.1.1 [ヒト医学機微データ]の取り扱いの指針.....	11
4.1.2 「機微情報を含まない研究データ」取り扱いの指針.....	12
4.1 公開共有の種類の変更の指針	14
4.2 利活用データの利用許諾ライセンスの指針.....	14
4.3 利活用データの保管期間とリポジトリでの管理方法.....	15
4.4 公共リポジトリの利用の指針	15
4.5 メタデータの付与に関する指針.....	15

4.6	データ管理者の転出対応に関する指針.....	16
-----	------------------------	----

1. 目的

国立研究開発法人理化学研究所（以下、研究所）バイオリソース研究センター（以下、BRC）は、「信頼性」「継続性」「先導性」を事業の柱として、研究コミュニティからのバイオリソースの寄託もしくは譲渡を受け、厳格な品質検査による真正性の確認、保存を行った後、それらを研究コミュニティへ提供する事業、加えてバイオリソースの利活用を促進するための研究開発を実施するセンターである。

BRC 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」を継承し、情報技術を通じた研究所の研究推進及び研究成果最大化に貢献する。BRC 研究データガイドライン（以下、本ガイドライン）は、BRC 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、BRC 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. 本ガイドラインの改訂方法について

本ガイドラインは、原則 1 年に一度の改訂検討を行う。改訂にあたっては、センター長の定めるガイドライン策定委員（会）が行い、センター会議あるいはそれに準じる会議体での検討を経て、センター長が承認する。

3. 本ガイドラインにおける用語の説明、定義

3.1 本ガイドラインにおける「データ」の定義

本ガイドラインにおける「データ」とは、電子化された情報を指す。従って、本ガイドラインは、ディスク等の媒体そのものの管理ではなく、媒体に記載された「データ」の活用に関する指針である。なお、研究著作として扱われる論文、およびソフトウェア等の計算機上で実行されるプログラム類は研究データに含まれない。（著作は著作権法の下にあり、ここに示す「データ」とは扱いが異なるため。）また、紙に書かれた情報は、媒体である紙とは切り離せないで、本ガイドラインでは対象外とし、その扱いについては、別途定められた指針（研究記録管理規程及び、バイオリソース研究センターにおけるラボノートブック取扱い要領等）を参照する。

- 「データ」の例
 - 電子ファイル
 - データベースに格納したデータ
- 「データ」ではない例
 - 著作であり、本ガイドライン以外のルールで扱われるもの
 - プログラム（バイナリファイル、ソースコード）
 - 論文
 - 紙の書類
 - ハードディスク

3.1.1 データ内容に基づく分類（区別）

本ガイドラインは、BRC で扱う「データ」のうち、「研究データ」のみを対象としている。本ガイドラインでの「研究データ」を明確に定義するために下記の通り分類（区別）した上で定義する。

データ内容に基づく分類（区別）

- [BRC 事業データ]
 - [研究データ]（本ガイドラインの対象）
 - [ヒト医学機微データ]
 - [機微データを含まない研究データ]

それぞれの詳細について、以下に示す。

- [BRC 事業データ]

BRC のバイオリソース事業で扱う全てのデータ。[BRC 事業データ]は、[研究データ]と[研究以外の事業データ]に区別される。
- [研究データ]（本ガイドラインの対象）

BRC 事業データのうち、生物、あるいはその派生物（遺伝子、遺伝情報、各種標本等）に関するデータを指す。研究活動によって生じたデータ及び、バイオリソースに直接的に付随する情報を含む。

[研究データ]は、[ヒト医学機微データ]と、[機微データを含まない研究データ]に区別される。

- [ヒト医学機微データ]
- 研究データのうち、ヒトを対象とする医学系研究における機微情報を指す。特別な扱いが必要であるため、次項目の[機微情報を含まない研究データ]とは区別される。[ヒト医学機微データ]には、個人識別符号が含まれる個人情報や要配慮個人情報（政府統一基準：機密性2情報、理研格付け区分「秘」に相当する情報が含まれる可能性がある。

例：

- ヒト iPS 細胞に関する個人情報
- ヒト個人のゲノムデータ

- [機微情報を含まない研究データ]
- 上記のヒト医学機微情報を含まない研究データを指す。具体的には、研究活動で得られたデータ、バイオリソースカタログ、特性データベース等のために取得、作成、計測したデータ等を指す。バイオリソース利用状況を統計的に管理した情報、文献情報など、バイオリソースに関連して利活用可能なデータも含まれる。特殊な場合として、バイオリソース利用において極めて重要な由来情報である寄託者の個人名も[機微情報を含まない研究データ]に含まれる。ただし、バイオリソース利用者情報など一般に個人情報として扱われる情報等は含まれない。

例：

- モデル生物を用いた実験データ
- バイオリソースカタログのために収集したデータ
 - リソースの基本情報
 - リソースの特性情報
 - リソースと寄託者名とのつながり
 - リソースと文献とのつながり

- [研究以外の事業データ]

BRC 事業データのうち、バイオリソース事業を行う上で取得された、個人や団体等の情報を指す。本ガイドラインでは、この研究以外の事業データは対象外とする。この[研究以外の事業データ]に関しては、別途策定されるガイドラインにて詳細を定める。

例：

- バイオリソース利用者の個人情報
- バイオリソース提供にかかる課金の情報

3.1.2 データのライフサイクル視点からの分類

研究データは、一次データ、二次データの生成に始まり、共有や保管のためにリポジトリへの登録、データ公開、データ削除に等々のライフサイクルを辿る。以下に、研究データのライフサイクル、利活用視点で分類（区別）を示す。

ライフサイクル視点による研究データの分類（区別）

- [センター内研究データ]
 - [一次研究データ]
 - [センター外由来一次データ]
 - [センター内産出一次データ]
 - [二次研究データ]

□ [センター内研究データ]

全ての研究データは広く利活用されるべきものだが、実際的には、[所内限定公開] [所内外限定公開] [公開]のいずれか（[公開共有の種類](#)を参照）に至るまでに段階を要する等の理由から、センター内（研究室を含む）における利活用に制限される場合がある。このような状態の[研究データ]を[センター内研究データ]と呼び、[センター外利活用研究データ]と区別される。[センター内研究データ]は、一定期間の後、[センター外利活用研究データ]となる可能性がある。

[センター内研究データ]は、データ処理・加工の観点から、さらに[一次研究データ]と[二次研究データ]に区別される。

□ [一次研究データ]

センター内研究データのうち、BRC にもたらされた未加工のデータを指す。[一次研究データ]は由来の観点から、[センター外由来一次データ]と[センター内産出一次データ]に区別される。

□ [センター外由来一次データ]

センター外から得られ、未処理・加工のデータを指す。処理・加工とは、計算処理その他によって、元の内容（テキスト、画素等）を改変することを指す。寄託者、外部機関の共同研究者から委譲を受けたデータには、MTA による利用制限がある場合がある（[利活用データの利用許諾ライセンスの指針](#)を参照）。

例：

- 寄託者から得られたバイオリオースの基本データ、特性データ（寄託者の名前そのものは、基本データ範疇とする）
- 外部データベースからインポートしたデータ
- 共同研究者から移譲を受けたデータ

□ [センター内産出一次データ]

センター内で得られ、未加工のもの。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、考案して作成した研究データ等が含まれる。

例：

- BRC で計測、取得したデータ
- BRC で計測した、Web アクセス解析のデータのうち、公開目的のもの

□ [二次研究データ]

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したもの。加工とは、計算処理その他によって、元の内容（テキスト、画素等）を改変することを指す。

データは加工を繰り返すことがあり、トレーサビリティの観点から元データとの関連付けを行い、管理するよう務める。

例（より詳しい説明）：

- センター外由来データを加工（センター内由来データが混ざる場合も含む）して得られたデータ
- BRC で計測、取得したデータさらに解析して得られたデータ。

□ [センター外利活用データ]

所属長により、センター外に公開（限定公開含む）し、利活用すると判断されたデータを指す。

例：

- バイオリソースカタログ（公開されたもの）
（事業色が強いが、生命研究に利活用される根本的なデータ（リソースの固有番号等）を含むため、利活用データと分類
- バイオリソース特性データベース（公開されたもの）
- コンソーシアムで共有／公開するデータ（限定付き公開を含む）

3.1.3 その他のデータの分類

3.1.3.1 [メタデータ]

他のデータを説明するために付け加えられたデータ。新たに作成され付与される場合と、既存のデータが他のデータを説明する役割を持って結合される場合がある。データとしての扱いは、メタデータもデータも同じであるが、「データには、それを説明するためのデータが必要である」という文脈で用いられる。

データをデータで説明する場合には、1)テキスト等で人間が読めるように記述する場合と、2) 機械可読、計算可能とするために、データに付与された ID 同士のつながりとして実装する場合がある。正確な情報提供や実用的なデータ検索を実現するためには、2)の方法が望ましいが、その前段階として1)も重要であるため、いずれかの方法を選択すればよい。

例：

- バイオリソースのメタデータ
 - リソースの生物種名と公共 ID、提供保存機関、提供元等（リソースの基本情報と同義）
 - リソースの生物学特性に関するデータ（リソース特性情報と同義）
 - 表現型データ、アノテーション結果
 - 改変された遺伝子、その公共 ID 等
- 文献のメタデータ
 - 出版日、著者、文献で使われたバイオリソース名や ID 等

3.2 その他関連する用語

3.2.1 [公開共有の種類]

基本指針では、公開共有の段階として「非公開」、「所内限定公開」、「所内外限定公開」、「公開」が定義されており、本ガイドラインもこれを継承する。

公開共有の種類は、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。

1. [非公開] [データ生成者]および[データ管理者]のみがアクセス可能な研究データ。
2. [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研究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。
3. [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]を対象にした研究データ。

3.2.2 [データ生成者]

研究データを生成した者をいう。個人のみでなく、グループや研究室を、この[データ生産者]と扱って良い。

3.2.3 [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役割を持つ個人をいう。

3.2.4 [データ利用者]

利活用データを受領し利用する者（個人あるいはグループ・団体）をいう。

3.2.5 [データリポジトリ]

複数のユーザーでデータを集積し、保存する場所

3.2.6 [(情報の) トレーサビリティ]

情報の流通経路を辿ることができること。ここでは特に、情報が何に由来するのかが明白になること、二次データを作成する際に、その由来となるデータが明らかになることで、データの由来を辿れることを指す。トレーサビリティの

確保は、科学データの証拠能力の確保のためにも重要であることから、BRC ではトレーサビリティの確保に積極的に努める。

4. BRC における研究データの取り扱い、管理の指針

4.1 研究データ取り扱いの指針

(全体指針) 3.1.1 に示した [研究データ]の管理には、[データ管理者]を指定した上で、原則として、1) 研究所が整備する研究データリポジトリ、あるいは2) 室・チームで管理するレポジトリあるいはデータベースを利用する。個人利用の PC 「のみ」におく場合は、所属長の許可を得る。リポジトリに登録された研究データは、[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、公開共有の種類¹⁾の指定、変更を所属長の承認を得て行う。そのため、開発室、チームは、[公開共有の種類]の指定、変更の具体的な方法を個別に定める。この方法には、レポジトリのフォルダに権限をつけ、公開共有の種類に従ってファイルを配置する等の方法が含まれる。

ただし、レポジトリの活用にあたっては、研究所が整備するレポジトリ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とに留意する。研究所にレポジトリ増強の要望を行う場合、客観的根拠が必要となることにも留意する。

室・チームで管理するレポジトリあるいはデータベース、個人の PC のみにおく場合は、当該の[データ管理者]が責任をもって管理する。それらデータをレ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。

4.1.1 [ヒト医学機微データ]の取り扱いの指針

理研では、人を対象とする医学系研究、あるいはバイオリソース整備の実施に当たり、理研内で生成、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を機微性が極めて高い要配慮情報として厳重に取り扱うこととなっている。一般的には、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

報等の取り扱い規程等の国立研究開発法人理化学研究所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。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必要がある。

具体的に BRC としては、「理研 BRC 臨床情報データ利用ガイドライン」¹に従って取り扱う。

4.1.2 「機微情報を含まない研究データ」取り扱いの指針

上記[ヒト医学機微データ]に含まれない研究データ、すなわち、「機微情報を含まない研究データ」に関しては、下記の指針に従い管理すること。

4.1.2.1 [一次データ]に関する作成および入手の指針

- 全体指針で定めたレポジトリに、[\[公開共有の種類\]](#)等を参照して、開発室、チーム毎に適切なユーザー権限を付与するか、あるいはルールを取り決めて保存する。
- 保存場所は、第三者にもわかるよう、ラボノート等に記録する。
- 共同研究契約や MTA(Material Transfer Agreement)が締結されている場合は、その契約との関連が明確になるよう管理する。

例：

- リソース番号と紐づく MTA データベースの運用
- データレポジトリのフォルダに、MTA との関連情報ファイルを置く

4.1.2.2 [一次データ]に関するデータ加工の指針

- 全体指針で定めたレポジトリに[\[公開共有の種類\]](#)等を参照して、開発室、チーム毎に適切なユーザー権限を付与するか、あるいはルールを取り決めて保存する。
- トレーサビリティを確保するため、由来するデータを明確に記録し、元データとの関連がわかるように保存するよう努める。由来の記録は、同じレポジトリ内にファイルを作成し記述する方法や、ラボノート等を活用して行う。

4.1.2.3 [センター内研究データ]から[センター外利活用データ]への移行の指針

- [センター内研究データ]から[センター外利活用データ]へ移行は、研究の責任者たる所属長の判断および承認により行う。研究データの管理、公開、共有及び利用に関する基本方針(理研データポリシー)ⁱⁱに従い、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

等が挙げられる。

- 全体指針で定めたレポジトリに[\[公開共有の種類\]](#)等を参照して、適切なユーザー権限を付与して保存する。公共レポジトリの利用に関しては、「[\[公共レポジトリ利用の指針\]](#)」に従う。
- 「[\[利活用データの利用許諾ライセンスの指針\]](#)」に従い、データの利用許諾に関するライセンスを付与する。
- 「[\[メタデータの付与に関する指針\]](#)」に従い、適切なメタデータの付与に務める。

4.1.2.4 データの削除の指針

- 開発室、チーム毎に、データ削除管理簿を作成する。
- 基本指針に従い、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。
- 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。

- 保管期限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リポジトリから削除するとともに、その理由をデータ削除管理簿に記録する。
- 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。
- 研究データの保管において、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。

4.1 公開共有の種類の変更の指針

「[研究データ取り扱いの指針](#)」に従って指定した[[公開共有の種類](#)]は、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。

4.2 利活用データの利用許諾ライセンスの指針

利活用データの公開にあたっては、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ライセンスを示すよう努める。

- 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
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 (CC) を利用する。
- 著作権者の権利保護に重点を置いて独自にライセンスを定める場合
共同研究契約や MTA(Material Transfer Agreement)が締結されている場合には、それに従ったライセンスを明示する。例えば、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等を行う。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。

4.3 利活用データの保管期間とリポジトリでの管理方法

基本指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。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が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。

4.4 公共リポジトリの利用の指針

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リポジトリに登録し公開するよう努める。

4.5 メタデータの付与に関する指針

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。BRC 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。研究データを産出した研究者は、BRC 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、所属長の了承の下でクリエイティブコモンズライセンス等の公共で用いられるライセンスを付し、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。

4.6 データ管理者の転出対応に関する指針

リポジトリに登録された利活用データの管理は、原則登録者あるいは登録者の所属長が責任を持って行う。登録者が BRC から転出する場合には、登録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、BRC の担当部署に相談し、その管理方法について決定する。

以上

脚注：

i

https://cell.brc.riken.jp/ja/hps/hps_diseaselist_index/clinical_data_user_guidelines/clinical_data_guidelines

ii <https://www.riken.jp/medialibrary/riken/about/info/datapolicy.pdf>