

# 報告書

2025年3月17日

## オープンサイエンス活動評価項目に関する提言

### 中間報告

理化学研究所

オープンライフサイエンスプラットフォーム

オープンサイエンス活動評価作業部会

理化学研究所 オープンライフサイエンスプラットフォーム  
オープンサイエンス活動評価作業部会 部会構成員名簿  
(2025年3月17日現在)

理化学研究所

大浪 修一	生命機能科学研究センター
奥村 嘉宏	脳神経科学研究センター
粕川 雄也	生命医科学研究センター
櫛田 達矢	バイオリソース研究センター
小林 紀郎	情報統合本部
二階堂 愛	生命機能科学研究センター
榎屋 啓志	バイオリソース研究センター
山縣 友紀	情報統合本部
山本 泰智	情報統合本部

九州大学

伊藤 隆司	医学研究院
-------	-------

農業・食品産業技術総合研究機構

川村 隆浩	農業情報研究センター
-------	------------

データサイエンス共同利用基盤施設

五斗 進	ライフサイエンス統合データベースセンター
------	----------------------

国立情報学研究所

込山 悠介	コンテンツ科学研究系
-------	------------

国立遺伝学研究所

藤澤 貴智	生命情報・DDBJセンター
-------	---------------

大阪大学

水口 賢司	蛋白質研究所
-------	--------

以上 15 名

## 要旨

先進の情報通信技術(ICT)の発展により、研究者間の連携が強化され、研究成果の共有や公開の機会が拡大している。このような環境のもと、研究の多様性を尊重しつつ互いの貢献を認め合いながら科学を発展させることが期待されている。本来このようなオープンサイエンスは、研究者自身がインセンティブを持って取り組み、成果が還元されるエコシステムとして機能すべきものである。しかし、実際にはその理想と大きな乖離がある。

本提言は、理化学研究所内外の研究者で構成されるオープンサイエンス活動評価作業部会が、生命科学分野におけるオープンサイエンスを推進するインセンティブの一つとして機能するような活動評価の項目を提言するものである。1章では背景の説明、2章では本提言が対象とするオープンサイエンスを「オープンデータサイエンス」として定義し、日本における現状と、研究者主導によるオープンデータサイエンスのあり方について述べる。3章ではオープンデータサイエンスにおいて中核となる「研究データ」の取り扱いについて、生命科学分野における論文との関係、データ共有や公開の利点とリスク、研究データの流通や持続性の観点から分析を行う。そして、これらを踏まえ、4章では具体的な評価項目を提言する。すなわち、1)良質な研究データの実現、2)データ流通、3)データ利活用のアウトリーチ活動、4)データ管理及び持続性、5)研究コミュニティへの貢献、6)人材育成、の6つの評価項目により、多面的かつ適切な評価を科学コミュニティで行っていくこと求めるものである。

本文書は中間報告であり、研究者などからのフィードバックを反映しながら、今後も評価項目の検討を進める。また、4章に掲げた評価項目に対する具体的な方法や定量的・定性的な値の定義については、最終報告において公開する予定である。

## 本提言への経緯

理研内で生命科学におけるオープンサイエンス活動を推進する研究プロジェクトであるオープンライフサイエンス(OLSP)において、オープンサイエンス活動評価基準を定めるべく、オープンサイエンス活動評価作業部会が設置された。

第一回目の会合は2021年8月に行われ、理研生命科学系のセンターに所属する研究者を含む10名の理研研究者と1名の外部研究者の合計11名により行われ、参加者によるブレインストーミングを行った。2021年11月1日に開催された第二回目の会合を行った後、外部研究者と意見交換を行うため、6名の外部研究者を招聘し、2022年1月18日にワークショップ（検討会）を開催して議論した。

さらに、2023年3月14日、2024年9月2日に外部研究者を招聘して検討会を行いつつ、随時SNSを利用した提言原稿の確認や数十回にわたるコアメンバーによる議論を行った。この過程で、活動評価指標の具体的な計測・算出方法に至るには長い時間にわたる議論が必要と判断されたことから、活動評価指標の提示までを中間報告として取りまとめることで文書作成が進められ、2025年3月17日にOLSP運営会議において中間報告の公開が承認された。

なお、最終報告については、最新の国内外のオープンサイエンスの動向、活動評価指標の具体的な計測・算出方法、国内外の研究者のフィードバック等を取りまとめながら提言内容を充実させ、公開する予定である。

## 1. 本提言の背景

科学の主要な目的は、自然界の法則を発見し理解することにある。これまで、人類が科学を通して得た知識は論文として記述・共有され、科学の発展に大きく寄与してきた。科学者は、自らの研究成果や新たな視点、解決に至る方法論などを論文として発表し、学术界と共有することで、さらなる知識の創出と議論の深化を促してきた。このように、科学で得られた知見は、原則として誰もがアクセスできる形で共有され、自由な発想による研究が促進されてきた。

学術雑誌の登場以来、論文を通じた研究成果の公開が一般的となり、科学はオープンなものと考えられるようになった。しかし、現実には研究データの共有に関する課題が存在する。具体的には、インセンティブの欠如、物理的・権利的な障壁、有料学術誌の購読費用などが、科学知識へのアクセスが制限する要因となっている。

近年のオープンサイエンスは、科学の知識共有という基本理念を維持しつつ、現代のメディアやインフラを活用してアクセス障壁を低減し、科学の質、研究の速度、信頼性を向上させることを目指している。

生命科学では、オープンサイエンスという概念が確立する以前から、研究データの共有が重視されてきた。これは、生命科学が生物種の多様性のみならず、研究手法においても多様なアプローチが求められる分野であるためである。研究データの生成には、独自の科学的問題設定や先進的な実験設計、データの解析手法が必要となり、研究者間の連携が不可欠である。また、オミックス技術の進展により、遺伝子、タンパク質、転写、代謝、表現型などの網羅的データが得られるようになったことが、データ駆動型科学の進展を促してきた。さらに、計測技術の急速な進展により、多種多様で大量のデータが生成されるようになり、それらを統合し、人工知能を活用することで、より複合的で高度な解析が可能となっている。

一方で、研究データの共有や公開には、多くの課題が伴う。データを利活用しやすい形式整理・変換する必要があり、オープン・アンド・クローズ戦略[7]の策定を通じて、公開すべきデータと制限すべきデータの管理が求められる。また、共有されたデータの適切な維持と管理が不可欠であり、継続的な運用が必要となる。さらにデータ安全保障の観点から外国為替及び外国貿易法（外為法）による規制や、軍事転用可能な技術に関連する管理といった新たなデータ管理規制にも対応しなければならない。

オープンデータの実現には、FAIR原則[5] (Findable:見つけられる, Accessible:アクセスできる, Interoperable:相互運用できる, Re-usable:再利用できる)が重要な基準となる。この原則を満たすためには、データ共有の標準化と利活用促進、データ形式の変換や再編纂、メ

タデータの付与、データの正確性確認、適切なデータライセンスの選定など、多岐に渡る作業が必要となる。しかし、これらの作業は、研究者にとって追加の負担となり、特に新たにオープンサイエンスに取り組もうとする研究者にとっては障壁なる。一方、現行の評価指標は、公開データ量やデータアクセス数といった定量的な基準が主であり、データ共有の努力を直接評価するものではないため、研究者にインセンティブとしては機能しにくい。このような課題を克服するためには、研究者に適切なインセンティブを与える新たな評価項目と基準が必要である。さらには、個人だけでなく組織レベルでの評価制度を整備し、オープンサイエンスの推進を支援する仕組みを構築することが求められる。

オープンサイエンス活動評価作業部会は、生命科学分野に焦点を絞り、オープンサイエンスの要素の1つであるオープンデータに重点をおいた活動の理想像を「オープンデータサイエンス」として定義することを試みるとともに、そのインセンティブの根拠となるオープンデータサイエンス活動評価の項目策定を研究者が自ら行うことで、この課題解決を試みることを目的として、2021年8月に組織、初会合が開催された。その後多くの対面あるいはオンラインでの議論を行い、2025年3月17日の取りまとめを行い、最終的に本提言を公表するに至った。

## 2. 本提言で扱うオープンサイエンス：オープンデータサイエンス

本提言の議論を明確にするため、対象とするオープンサイエンスの範囲を定義する。研究活動は多面的であり、オープンサイエンスの概念も多様であるため、すべての側面を一律に実現することは困難である。そのため、国や組織、分野ごとに異なる事情を踏まえて定義されるべきである。

本提言では、議論の範囲を生命科学分野に限定し、日本国内の状況に即した定義を行った上で、評価項目について検討する。この方針は国際的な議論を妨げるものではなく、今後の国際的な研究コミュニティの多様な事情を考慮しながら詳細な検討を行うことを前提とする。

### 2.1. 日本および世界における現状

オープンサイエンスの本格的な議論は、2013年6月に英国にて行われたG8科学大臣会合において、論文のオープンアクセス化と研究データの公開が提言されたことを契機に始まった[2]。日本においても、公的研究資金による研究成果（論文、研究データなど）を、科学界のみならず社会全体がアクセス・利用可能とする新たな科学研究の進め方について議論がなされ、2015年3月に内閣府より「国際的動向を踏まえたオープンサイエンスに関する検討会」の報告書が公表された[3]。以降、日本政府や日本学術会議等などが科学政策

の一環としてオープンサイエンスを推進し、様々な戦略が提示されてきた。

一方で、研究データの管理や共有の手段として、データ種や研究分野ごとに特化した公共リポジトリが運営されている。論文投稿の条件として、研究データの公共リポジトリへの寄託が求められる場合もある。さらに、論文雑誌出版社などが論文だけでなく、関連データの収集・管理を進めており、商業的組織による研究データの独占の懸念も指摘されている。

このように、現在のオープンサイエンスは、政府や商業的組織が先導する形で進められており、研究者のリーダーシップや自主性が十分に発揮されているとは言い難い状況にある。

## 2.2. 研究者主導のオープンサイエンスの理想像

オープンサイエンスは、科学研究をより効率的かつ多様に発展させるための戦略の一つであり、研究者自身が主体的に推進すべきものである。そして、オープンサイエンスを実践する研究者には、その貢献が自身にも還元さえるようなエコシステムの構築が求められる。さらに研究者間の連携が深まり互恵関係が築かれることが望ましい。

得意、生命科学研究は以下のような特徴がある。

- ・生物種の多様性
- ・研究手法（アプローチ）の多様性
- ・データ網羅性の重視（オミックスデータなど）
- ・計測機器の多様性
- ・研究データの多様性と膨大なデータ量
- ・応用研究の多様性（農業、材料・素材、創薬、医科学など）

近年の情報通信技術(ICT)や知能情報科学等の情報科学の進展により、従来の研究方法にICTを取り入れ、高効率に多様で膨大な研究データを共有することが求められている。さらに人工知能の活用やデータ駆動型研究の推進による研究を変革は喫緊の課題である。

また、研究者には、研究成果が適切な方法で得られたことを客観的に示し、生命科学研究コミュニティのみならず、社会全体への還元を推進することも求められる。

本提言では、オープンサイエンスを、データサイエンスの視点で「オープンデータサイエンス」としてとらえ、生命科学の現状等を踏まえた具体的な議論を展開する。

## 2.3. 生命科学の現状を踏まえたオープンデータサイエンス

以上をふまえ、生命科学分野を対象としつつ、可能な限り一般化したオープンサイエンスの定義を示す。

オープンデータサイエンスの定義:

オープンデータサイエンスとは、先進の情報通信技術(ICT)を活用し、研究者間の連携を

強化しながら研究成果を共有・公開することで、研究の多様性を尊重しつつ発展的に科学を推進していくことを指す。

この定義は、最先端の ICT 技術を活用して科学の可能性を最大限に引き出すことを前提としており、オープンサイエンスを単なる手法ではなく「科学そのもの」として捉えている。

オープンデータサイエンスの実現のための戦略

オープンデータサイエンスを実現するために、以下の戦略が求められる。

- (1) 研究活動を加速し、研究の高度化・多様化に迅速に対応する
- (2) 研究成果を社会に還元し、共同研究や応用研究を通じた産学連携を促進する
- (3) 長期的な視点を持ち、透明性・再現性・持続性のある研究成果の共有を推進する

ICT の導入による研究手法の変革に加え、研究成果の信頼性を確保しながら、研究コミュニティや社会への還元を積極的に進めることが研究者に求められる。また、研究者自身の研究成果のみならず、オープンサイエンスへの貢献そのものが評価される仕組みを構築する必要がある。

### 3. 生命科学における研究データの取り扱い

本章では、議論の対象となる研究データの範囲を明確に定義する。まず、本提言では研究記録管理に関する詳細な議論を行わないが、研究データは論文に疑義や指摘が生じた際に対応できるよう適切に管理されるべきであり、研究者はこれを長期的に管理する責任があると考えられる。

一方、本提言では、オープンサイエンスの推進に資する研究データに焦点を当てる。すなわち、科学の発展のために流通、利活用されるデータやデータセットを研究データの範囲と定める。

#### 3.1. 研究論文と研究データ

生命科学における論文と研究データの間を A. 実験(ウェット)研究者とデータ(ドライ)研究者、B. ICT 技術や計測技術の進展 の 2 つの視点で議論する。

##### A. 実験研究者とデータ研究者

データ研究者（ドライ研究者）は、コンピュータを用いたデータ解析やシミュレーションを通じて生命現象の解明を目指す。データやプログラムを扱うため、研究データの収集や生成したデータ、プログラムの公開が比較的容易である。



一方、実験研究者は試薬やバイオサンプルを用いて、生命現象を探究する。データとは対極にあると思われがちだが、実際には研究の課題設定やその解決手法の探求において重要な役割を果たし、その結果として研究データが生み出される。そのため、実験の詳細を記述した論文やメタデータとともに研究データを共有・公開することは、生命科学全体の発展に大きく貢献する。

研究データの共有・公開を促進する具体策として、計測機器の設定や条件を自動的に出力する機能の導入や、データの流通障壁を低下させるためのデータ形式の標準化、研究データの利活用を容易にする関連ツール類の提供が有効である。

## B. ICT 技術や計測技術の進展

研究論文は通常、導入 (introduction)、方法 (method)、結果 (results)、議論 (discussion) から構成される。最も客観的に記述されて普遍的な価値を持つのは「結果 (results)」である。研究データも客観的なものであり、本来は結果に記述されるべきだが、近年の計測技術の進展により、大規模かつ多様なデータが生成されるようになった。これらをすべてを論文に記述することは現実的ではない。そのため、論文執筆者は結果の論拠となるデータセットを作成し、論文誌と同時に投稿、またはデータリポジトリに配備して公開するなどの方法で公開を行うことが一般的になっている。

### 3.2. 研究データ共有・公開に関する研究者の利点とリスク

研究データの共有・公開にあたっては、個人情報や機密情報の管理、研究に関わる契約条件を考慮し、適切なオープン・アンド・クローズ戦略を策定することが重要である。ただ単にデータをオープンにすればよいわけではなく、リスクと利点を慎重に検討した上で、適切な戦略に基づいてデータを扱う必要がある。

研究データを共有・公開するリスクとして、以下が挙げられる。

#### リスク 1. 誤った利活用

研究データの改ざんや誤った解釈、不適切な解析プログラムの適用、不正な目的での利用により、研究者や研究分野の信頼性を損われるリスクがある。

#### リスク 2. 知的財産権の問題

研究データセット (データベース) が持つ著作権を正しく扱わなかった場合、原著者に損害を与え、研究者自身の信頼を失うリスクがある。

#### リスク 3. 研究者自身の競争力の低下

研究データを公開することで、他の研究者が同じデータを利用して新たな研究を行い、研究者自身の研究競争力が低下するリスクがある。

一方、利点として、以下が挙げられる。

#### 利点 1. 研究の透明性と信頼性の向上

公開された研究データにより、研究結果を再現・検証が可能になり、研究の質や信頼性、研究者の評価が向上する。

#### 利点 2. 研究引用の増大

他の研究者が公開データを引用することで、研究者の影響力が増し、研究の方向性決定や標準化、研究資金、人材獲得などにおいてリーダーシップを獲得できる。

#### 利点 3. 新たな研究・イノベーションの創出

公開データを他の研究者が利用することで、新たな研究やイノベーションが生まれ、科学の発展につながる。

#### 利点 4. 研究データの再利用と研究効率の向上

研究者同士がデータを共有することで、重複実験を削減し、研究の効率化が進む。

以上のリスクと利点を踏まえ、研究現場では多角的な視点からオープン・アンド・クローズ戦略を策定する必要がある。この戦略立案は難易度が高く、研究者には高い意識と正確な判断力が求められる。また、戦略立案そのものもオープンサイエンス活動の重要な一環として評価されるべきである。

### 3.3. 研究活動と研究データ

研究データは研究活動の成果であり、両者は極めて密接に関係している。本節では、生命科学の研究活動を6つの視点から議論し、オープンサイエンス活動の評価基準の定義につなげる。

#### 3.3.1. 良質な研究データ

質の高い研究データの作成には相応のコストがかかる。一方、利活用の観点からは、\*\*データの「価値」\*\*が重要となる。研究データの品質や価値を適切に評価できれば、良質なデータ作成のインセンティブとなり得る。

データそのものの質や価値は、科学的・社会的ニーズ、独自性、希少性などに依存するが、それらを絶対的な基準で評価することは困難である。

しかし、オープンサイエンスの推進には、良質なデータ作成のインセンティブを生じさせるこれらの評価が重要であることは指摘しておく必要がある。また、オミックス研究の進展により、各トピックにおけるデータの網羅性は重要な評価軸である。さらに、データのアクセス数などの利用実績も評価軸として多く用いられている。

オープンサイエンスにおいては、データ形式や記述方法の標準化が利活用性の向上につながる。例えば、国際標準に従ったデータ形式の採用や、統制語彙を用いた記述によってデ

ータの相互運用性が向上し、利活用が推進される。また、データを適切に活用できるよう、読み込み、表示、解析を容易であることも重要な要素である。商用ソフトウェアに依存するデータ形式は、データの質や情報量に影響を与えない範囲で、標準的な形式に変換することで流通を容易にできる。

また、メタデータの適切な付与によって、データの品質や価値は大きく向上する。研究データの管理においては、研究内容をよく理解したうえで、必要十分なメタデータの項目、利活用しやすい形式等を考慮してメタデータの管理を行っていくことが重要である。さらに、データ計測の条件として、作成日時・作成者や地理情報などを明記し、データ生成過程を記録することで、データの信頼性が高まる。バージョン情報や更新履歴の管理も、利活用を促進する要素となる。これらを記述するメタデータは、データそのものと同様に標準化された語彙やスキーマ、データモデルを採用することが望ましい。また、データセットのデータ数とデータ量が含まれるメタデータを付与すると、利活用の際に便利である。さらに、AIによるメタデータ記述が普及することを考慮し、手作業で付与されたものか、AIによるものかを明示することが、メタデータの正確性評価において重要となる。価値の高いデータの作成には、適切なコスト意識が必要である。しかし、現状、我が国ではこの認識が十分に根付いていない。研究データへのメタデータやアノテーション付与は、研究者自身が行うべき作業であり、そのコストに見合った評価を行うことが、結果としてデータの品質や価値向上につながると考えられる。

### 3.3.2. 研究データの流通

研究データの共有・公開は、流通しやすく利用しやすい形式や方法で行われることが望ましい。そのための重要な指針の一つとして、FAIR原則が挙げられる。

まず、データやメタデータをウェブ上でアクセス可能とするには、適切な識別子(ウェブアクセス可能な一意のID)を付与する必要がある。また、データやメタデータを検索し取得するためのウェブサイトの設置、プログラムを通じた利用を可能にするアプリケーション・プログラミング・インターフェイス(API)等の整備により研究データの利便性が向上し、流通が促進される。また、研究データやメタデータの共有・公開においては、オープン・アンド・クローズ戦略に基づき、適切なデータアクセス権や利用権を設定することが重要である。まず、データの\*\*知的財産権(知財権)\*\*などの権利情報を明確にし、それを利用者に提示した上で、データやメタデータへのアクセスを許可すべきである。このライセンス情報の提示方法としては、ウェブページやAPIを通じた表示のほか、メタデータにライセンス情報を直接記載する方法も考えられる。

さらに、認証機能を活用することで、データやメタデータの取得や検索ができるユーザーを適切に管理することも重要である。認証されたユーザーごとにアクセス権を調整し、必

要に応じて制限を設けることで、データの公開範囲を適切に設定できる。これにより、機密性を保持しつつ、研究データの効果的な共有・公開を実現することが可能となる。

### 3.3.3. データ利活用のアウトリーチ活動

研究データの利活用を推進するためには、データが共有・公開されていることを周知することも重要である。従来の広報手段としては、学会等での発表、論文での発表、プレスリリース等などが一般的である。これに加え、データやその利用方法を紹介するウェブサイトや動画コンテンツの公開も有効な手段となる。また、講習会や説明会といった対話形式での広報では、ユーザーのニーズにあった広報が可能となる。さらに、ユーザー対応窓口（問い合わせ先）を設置することで、直接的にフィードバックを得る機会とすることができる。研究データの公開には、データの編纂・整理、メタデータ・アノテーション付与、データの品質管理、データ公開の持続性の確保など、多くの手間やコストを要する。そのため、研究者がデータを積極的に公開できるよう、これらの負担を上回るインセンティブを提供する仕組みが必要となる。研究データの利用実績を研究者が認識できるようにし、それを研究成果の評価に反映させる仕組みを整備することが挙げられる。また、研究データの利活用を促進するためには、データの解析方法を研究者間で共有することや、公開されたデータを用いた研究成果を発表する際に、論文内に研究データとその利活用方法を明記することなど、新たな論文発表の文化を醸成することも重要である。

さらに、分野や社会での研究データの活用

も重要なアウトリーチ活動の一環である。生命科学分野は特に応用研究の幅が広く、農業、材料・素材、創薬、医科学など、さまざまな分野でのデータ活用が期待される。これらの分野に限らず、異分野での生命科学研究データの利活用は、新たな科学的価値の創出につながる可能性がある。また、産学連携を通じた研究データの社会還流は、産業界の発展に貢献するだけでなく、学術機関においても経済活動に結び付ける機会となり、研究資金の獲得や研究力向上につながる。加えて、データ利活用は、いわゆるシチズンサイエンス（一般市民が参加する科学活動）への対応という観点からも重要である。シチズンサイエンスの対象には、研究を職業としない一般の人々、将来研究者を目指す若手、教育目的でデータを活用する生徒や学生が含まれ、こうしたユーザーへの広報は、研究者への広報とは異なり、分かりやすさや丁寧な説明が求められるため、必研究者には特別の配慮が必要となる。

このように、研究データの利活用を進めるためには、適切な広報活動とインセンティブの整備が不可欠であり、研究者、産業界、社会全体が協力してデータ共有の価値を高めていくことが重要である。

#### 3.3.4. 研究データ管理及び永続性

生命科学では、日々膨大な量のデータが作成されている。研究者や研究機関には、これらのデータを研究者あるいは研究機関の知財として適切に管理し、活用できるよう計画を立案することが求められ。具体的には、オープンデータとして共有、公開するデータ、研究公正の証拠として保存するデータ、非公開として保持するデータなど、それぞれの性質に応じた適切な管理が必要である。また、永続的、すなわち長期的に研究データが安定、安全に共有、公開ができる体制の構築も重要であり、そのための継続的な取り組みが求められる。

研究データの永続的な共有・公開を実現するためには、国際的に恒久的な識別子（DOI: Digital Object Identifier）を付与し、リンク切れを防ぐことが有効な方法の一つである。しかし、データの長期保存には、研究者個人のみならず、所属機関や研究コミュニティ全体の努力も不可欠である。例えば、研究プロジェクトの予算終了や、研究者の異動によってデータの管理が困難になることがある。このような課題を克服し、長期的なデータ共有を実現している研究者や、データの永続性を保証する研究コミュニティや機関はオープンサイエンス活動として評価されるべきである。また、データの永続性を確保するためには、セキュアなデータ管理環境を構築または利用して、適切な認証とアクセス権を設定することも重要である。研究機関では、このような機能を備えた研究データ管理基盤の整備が不可欠となる。その基盤としては、リポジトリやネットワーク等の情報インフラの整備が基本となるが、加えてデータ管理を支援するミドルウェアやツール類の開発も必要である。研究分野や研究手法によって、毎必要とされる管理基盤が異なるため、それぞれのニーズに対応した研究者との協働による基盤開発が求められている。さらに、計画的なデータ管理を推進するためには、研究データ管理計画（DMP: Data Management Plan）の策定と運用が有効である。近年では、多くの研究プロジェクトでDMPの立案が求められ、研究費助成においてその策定を義務付けるケースも増えている。DMPを適切に運用することで、研究データの管理体制を強化し、長期的な利活用を可能にする基盤を整備することができる。

研究データの適切な管理と永続性の確保は、研究の信頼性や再現性を向上させるだけでなく、データの価値を最大限に引き出し、将来の研究活動にも貢献する重要な取り組みである。

#### 3.3.5. 研究コミュニティへの参加

研究データの管理や標準化、関連ツール類の開発・共有は、多くの研究者が集まり議論できる研究コミュニティの活動によって支えられている。研究者が、自らの研究成果の先進性を示すために、独自のデータの形式や仕様を採用することが考えられるが、可能な限

り広く認められている標準仕様を採用し、より多くの研究者がデータの利用しやすい環境を整えるである。研究の発展に応じて、既存の標準仕様との互換性を保ちながら仕様の更新を行うことも研究コミュニティの活動の一部としてで行うべであり、このような活動への貢献は、オープンサイエンスにおいては評価されるべきである。具体的な活動としては、国際コンソーシアム等における研究データや情報基盤に関する標準化、共通化、設計、実装、ルール策定に参加することが挙げられる。また、研究データを扱う解析プログラムやデータ管理ツールの公開やオープンソースツールの更新・保守など、ツールの共有と公開も評価されるべき活動である。さらに、安全かつ安心して研究データを活用できるように、研究データ管理におけるコンプライアンス（倫理や規範の順守）の強化することも重要である。これらに関連するツールや情報基盤の開発、構築、運用、共通化、標準化に関する活動もオープンサイエンスの一環としての評価対象である。また、研究データを介した連携にとどまらず、研究者同士の議論や論文を通じた知見・技術の交流もオープンサイエンスを進めるために不可欠な活動である。異分野間の連携が進むことも重要だが、同じ分野で活動している研究者や技術者と連携し、共に研究基盤を深化させることも大切な活動と認識されるべきである。

### 3.3.6. 人材育成

オープンサイエンスにおいては人材育成の重要性は高く、多様な専門知識と技術を持つ人材の確保や育成が、研究や社会への貢献、さらには持続的な発展の鍵を握る。そのような人材育成の目的として、まず、研究データの作成、編纂、維持、管理、共有、公開に必要な能力を習得し、それを共有することが挙げられる。さらに、研究データ基盤やインフラの構築・運用に必要な技術の習得と共有も重要である。関連する教育コンテンツ（教材）の整備や、その運用として、研修や学習の機会を提供していくことも求められる。加えて、新たな情報通信技術（ICT 技術）や研究手法の習得を通じて、研究の発展に貢献できる人材を育成することが求められる。オープンサイエンスの持続的な発展には、こうした人材の確保が不可欠であり、国際的な研究コミュニティにおいて競争力を高め、科学全体の発展に寄与できるリーダーとなる研究者を育てることも重要である。

これらの課題を実現するためには、指導的立場にある研究者や学習意欲の高い研究者だけでなく、組織としての積極的な関与が必要となる。例えば、組織によるオープンサイエンスに関する教育コンテンツの作成・公開や、講習会・勉強会の実施は、人材育成に有効な手段となる。さらに、組織として人材育成への貢献を評価し、本提言の項目を人事評価に組み込むことで、オープンサイエンスへの参加を促すインセンティブの向上も期待される。

近年ではデータ駆動型科学の進展に伴い、「データスチュアート」と呼ばれる職種の重要性が高まっている。データスチュアートは、データ管理や品質保証、アクセス権の制御など、データに関する専門的な役割を担う職種である。具体的には、データガバナンスの実施やデータの正確性や一貫性の維持、ライフサイクル管理、プライバシーやセキュリティの確保、データの使用に関するポリシーの策定や、データ利用者への支援が含まれる。こうした専門人材の育成は、オープンサイエンスの推進において欠かせない要素であり、今後さらに注力すべき課題の一つである。

#### 4. 評価項目の詳細

特にデータサイエンスの側面が強い生命科学研究により深く関連し、喫緊に評価項目の確立が求められるものは、オープンデータに対するものである。そこで、それ以外のオープンアクセス等の各項や、上述の人材育成を含むオープンサイエンス活動についても留意しつつ、研究成果の共有・公開に直接関連するオープンデータを主軸に据えて評価項目の列挙を行う。

さらに、オープンデータにより達成されてデータ集約型科学において、研究データ等が満たすべき原則を述べた FAIR 原則にも留意し、識別子、メタデータ、プロトコル、語彙、ライセンス等についても評価項目に含むよう定義する。

なお、最終的に評価に必要とされるものは、評価項目ごとに定められる評価指標である。評価指標は、透明性確保の点で評価の根拠が示され、さらに点数に換算されるよう定義され、広く公開されることを前提とする。本中間報告の範囲は、評価項目の列挙に留めることとするが、最終報告においては、評価指標としての点数の算出方法も含め、国内外の科学界の状況に合わせて適切に定め、さらには評価指標の運用を通じて検討、最適化していくこととする。

本提言では、3.3.節での議論を踏まえ、6つの評価軸

- (1) データの価値向上に関わる活動に関する項目
- (2) データ形式・流通に関する項目
- (3) データ利活用のアウトリーチ活動に関する項目
- (4) データ管理に関する項目
- (5) 研究コミュニティへの貢献に関する項目
- (6) 人材育成

を掲げて分類し、これまでに議論を踏まえて各評価項目について詳細に列挙する。なお、<>により付される記号は、対応する FAIR Metrics[6]の項（補遺1）を示している。

##### 4.1. 良質な研究データの実現に関する評価項目

科学的見地からデータの価値を、その生成過程を含め透明性をもって評価できるようにする。特に、データの価値向上に関する活動や付加価値を与えている活動に関する評価項目を列挙する。

[研究データの科学的、社会的ニーズとの対応に関する項目]

- (1.a) 対象とするデータの独自性、希少性、網羅性
- (1.b) 研究データのアクセス数等の利用実績
- (1.c) 研究データの形式が標準的なものを採用しているかどうか、また標準的な統制語彙を用いた記述がされているかどうか

[データの品質や価値を高めるメタデータに関する項目]

- (1.d) 研究データの計測の諸条件に関する情報や、(地理情報、作成日(時)、作成者等含む) データ生成過程等の由来が記述されているか <F2>
- (1.e) 研究データの形式を正しく記述されているか <I1, I2>
- (1.f) 研究データのバージョン、更新日時等の更新履歴に関する記述がなされているか <R1.2>
- (1.g) メタデータの形式について、標準化・統制されている適切な語彙、スキーマ、データモデルなものを採用しているかどうか。<I1, I2>
- (1.h) メタデータの正確性に関する情報記述されているかどうか。例えば、メタデータをキュレーションで付与しているか、AIで付与しているか
- (1.i) 研究データセットのデータ数とデータ量を記述しているか

#### 4.2. データ流通に関する項目

データ共有・公開において、データの外形的特質について満たすべき属性や量的評価軸を掲げ、具体的な評価を行うものである。FAIR原則を満たすデータ指標や、データ数、データ量、データにアクセスするためのプロトコルに関する項目がここに含まれる。

- (2.a) データあるいはメタデータにウェブアクセス可能な識別子(アクセッション番号)がついているか<A1+F1>
- (2.b) データあるいは、メタデータの取得および検索のためのAPIを含むインターフェースが提供されているか<F4><A1.1>
- (2.c) データ共有・公開のライセンスが明示されているか、権利等、知財権に関する情報が記述されているか(メタデータがあるか) <R1.1>
- (2.d) 認証(認証不要を含む)とアクセス権を適切に運用しているか <A1.2>

#### 4.3. データ利活用のアウトリーチ活動に関する項目



共有・公開したデータが利活用されるよう行われる広報活動に関する項目を列挙する。

- (3.a) データあるいはメタデータについての学会発表、論文発表、プレスリリース
- (3.b) データあるいはメタデータの紹介、利用方法説明のウェブサイト、動画等の公開、講習会やデータ説明会の実施
- (3.c) データを利活用したことを論文等で明記することを促進する活動
- (3.d) ユーザー対応窓口（問い合わせ先）の設置
- (3.e) 医学農学を含む産業界でのデータ活用、連携等の促進
- (3.f) 生徒・学生、非研究者を含む一般社会・研究後継者への広報

#### 4.4. データ管理及び永続性に関する項目

データ管理コンプライアンスに関する項目を列挙する。

- (4.a) データ永続性（サステナビリティ）の確保として、
  - ・ 研究データに付与されている識別子の永続性
  - ・ データを永続的（＝長期的）に共有・公開するための体制があるかどうか
- (4.b) 研究データ管理における、セキュアな管理環境の利用、および適切な認証とアクセス権の確保
- (4.c) データ管理計画（DMP）の作成、定期的見直し、および遵守

#### 4.5. 研究コミュニティへの貢献に関する項目

広く研究データを流通させるための枠組やガイドライン等の立案や、ツールやシステムとしての実装等の貢献に関する項目を列挙する。

- (5.a) 国際コンソーシアム等における研究データや情報基盤に関する標準化、共通化、設計、実装、ルール策定等を含む連携活動への参加
- (5.b) オープンサイエンスを推進するための、オープンソースのデータ管理ツール等の共有、公開、更新、および保守
- (5.c) データ管理コンプライアンスを向上させるための、ツール、インフラの開発、構築、運用、共通化、標準化の促進
- (5.d) 自らの研究・開発と類似した活動を行っている他研究者・技術者との、オープンサイエンス推進のための連携

#### 4.6. 人材育成

オープンサイエンスの永続的な推進や発展に必須となる人材育成に関する項目を列挙する。

- (6.a) 研究データの作成、編纂、維持、管理、共有、公開を行う人材の育成
- (6.b) 研究データの基盤等のインフラを構築、運用を行う人材の育成
- (6.c) オープンサイエンスに関わる教育コンテンツ（教材）の作成、公開
- (6.d) オープンサイエンスに関わる人材育成に資する講習会や勉強会の実施
- (6.e) 新しい技術や動向の調査と研究現場への導入

## 5. 議論とまとめ

本提言で掲げた評価項目については、国内外を問わず、より多くの研究者の間でコンセンサスを取りながら議論を進め、中間報告としてとりまとめられたものである。最終的には中間評価に対する研究者等のフィードバックを受け、さらに評価項目に対する具体的な方法や定量的・定性的な値の定義を行ったうえでオープンサイエンス活動評価ガイドラインとしてまとめる計画である。我々のガイドラインは、研究機関、大学等で採用されることで標準的で権威ある研究評価の方法として定着することを筆者らは望んでいる。このことを通じて、より多くの研究者がオープンサイエンスに興味を持ち参加することで、オープンサイエンスの発展につながることを切に希望する。

以上

## 参照

- [1] Jeroen Bosman, and Bianca Kramer. Defining Open Science Definitions.  
<https://im2punt0.wordpress.com/2017/03/27/defining-open-science-definitions/>
- [2] G8 Science Ministers Statement (2013 年 6 月 13 日)  
<https://www.gov.uk/government/news/g8-science-ministers-statement>
- [3] 内閣府 国際的動向を踏まえたオープンサイエンスに関する検討会  
<https://www8.cao.go.jp/cstp/sonota/openscience/>
- [4] openscienceASAP. Was ist Open Science?  
<http://openscienceasap.org/open-science/>
- [5] FORCE11, The FAIR Data Principles.  
<https://force11.org/info/the-fair-data-principles/>
- [6] FAIR Metrics  
<https://fairsharing.org/search?fairsharingRegistry=Standard&recordType=metric&page=1>
- [7] 内閣府 第 5 期科学技術基本計画 本文  
<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>

## 補遺 1. FAIR Metrics (日本語訳)

### 発見可能性 (Findable)

- F1. (メタ)データがグローバルに一意で永続的な ID を有すること
- F2. データがメタデータによって十分に記述されていること
- F3. データセットに対して、グローバル、ユニーク、かつ永続的な識別子をメタデータで記述すること
- F4. メタデータが検索・リスト可能であること

### 到達性 (Accessible)

- A1. 標準化された通信プロトコルを使っている
  - A1.1. プロトコルは公開されており、無料で、実装に制限がないこと
  - A1.2. 必要な場合は、認証や権限付与の方法を提供できること
- A2. データが利用不可能となったとしても、メタデータにはアクセスできること

### 相互運用性 (Interoperable)

- I1. (メタ)データの知識表現のため、定型、到達可能、広く適用可能な記述言語を使用していること
- I2. (メタ)データが FAIR 原則に従う語彙を使用していること
- I3. (メタ)データは、他の(メタ)データへの特定可能な参照情報を含んでいること

### 再利用性 (Reusable)

- R1. (メタ)データが、正確な関連情報を豊富に持つこと
  - R1.1. 明確でアクセス可能なライセンスと共に公開されていること
  - R1.2. データが来歴とつながっていること
  - R1.3. (メタ)データが、分野ごとのコミュニティの標準を満たしていること