

画像データの共有がもたらす生命科学の発展

— 理研のデータ共有エコシステムが世界をつなぐ —

概要

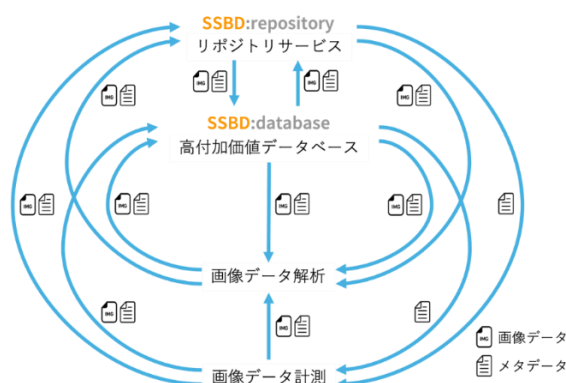
理化学研究所（理研）生命機能科学研究センター発生動態研究チームの大浪修一チームリーダー（情報統合本部生命科学データ共有開発ユニットユニットリーダー）、京田耕司技師、糸賀裕弥技師、情報統合本部生命科学データ共有開発ユニットの山縣友紀研究員（バイオリソース研究センター統合情報開発室研究員）らの研究チームは、生命科学分野の画像データの共有と再利用を促進する公共リポジトリ^[1]および高付加価値データベースを開発・公開しました。

本データベースは、画像データの標準化・共有促進を目指す国際コンソーシアム「foundingGIDE^[2]」の中核を担うエコシステムの一つと位置付けられており、科学の透明性と信頼性を高め、オープンサイエンスの進展を通じて社会全体における科学の発展に貢献すると期待されます。

研究から生み出されたデータの共有は科学の発展に不可欠ですが、論文発表時には迅速なデータ公開が求められる一方、豊富なメタデータ^[3]を付与するための時間やリソースが不足し、研究者はそのジレンマに直面していました。

今回、研究チームは、画像データ公開と共有の柔軟性を向上させるため、これまでチームが開発してきた画像データと画像解析データのデータベース「SSBD^[4]」を再構築し、公共リポジトリサービス SSBD:repository と高付加価値データベース SSBD:database の2階層のデータリソースを提供しました。

本研究は、科学雑誌『*Nucleic Acids Research*』のオンライン版「Database Issue（データベース特集号）」（10月31日付：日本時間10月31日）に掲載されます。



生命科学分野の画像データ共有のエコシステム

背景

近年、生体を対象とする画像取得技術が急速に進展しています。生み出される膨大な生物画像データを研究者が相互に参照・利用するためには、研究成果を共有する適切なデータベースが必須です。FAIR原則⁶⁾に従ったデータ共有は、データの発見可能性、アクセス可能性、相互運用性、再利用可能性を確保し、科学の発展に極めて重要です。特に共有されたデータの再利用は、新たな知見を生み出し、研究の加速に寄与します。また、研究者が膨大な量のデータを再利用できるようにするためには、データの意味や文脈を正確に理解した上で付与されたメタデータが不可欠です。しかし、論文発表時には迅速なデータ公開が求められる一方で、データを共有するためのメタデータを豊富に付与するためには、時間や人的リソースが必要で、その作業には高いコストが伴います。このため、発表直後にデータを公開する際に十分なメタデータを付与することが難しく、研究者は迅速な公開と、質の高いメタデータ付与の間でジレンマに陥っていました（図1）。

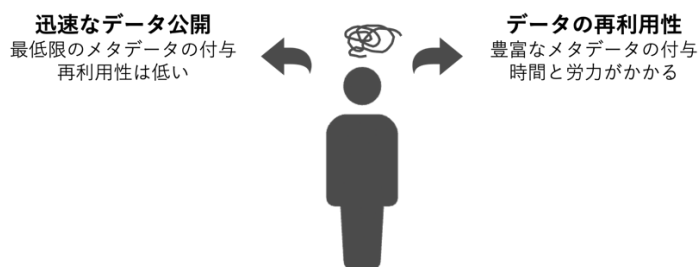


図1 データ共有時のジレンマ

研究者は論文発表直後にデータを公開する際、十分なメタデータを付与することが難しく、迅速な公開と質の高いメタデータ付与の間でジレンマに陥っていた。

研究チームはこれまでに、生命科学分野の画像データと画像解析により計測される定量データを共有、再利用するためのデータベース「SSBD」を開発してきました。SSBDは公開当初、論文で発表された画像解析後のセグメンテーションデータ⁶⁾やトラッキングデータ⁷⁾を主として、その元となる画像データを共有する公共データベースでした。2015年より、画像解析技術の発展やデータを介した最先端イメージング技術の開放を主な目的として、画像データのみを共有を開始しました。さらに、2016年より、著者からの要望に応えるかたちで、論文公開前に画像データを共有するリポジトリサービスを試験的に提供してきました。しかしながら、このサービスは正式な運用体制が整っておらず、機能やサポート面での課題が残っていました。また、データの迅速かつ安全な共有を実現するためには、正式なりポジトリサービスの構築と運用の改善も必要でした。

研究手法と成果

研究チームは、SSBD を再構築し、2 階層のデータリソースを提供することで、データ公開と共有の柔軟性を向上させました（図 2）。まず、1 階層目として公共リポジトリサービス SSBD:repository を整備し、学術誌上で発表された全てのデータを迅速に公開できるようにしました。迅速なデータ公開を可能にするために、SSBD:repository では、著者が、問い合わせ先やライセンス、生物種など最低限の必須なメタデータを記述するだけで、データを共有することができます。さらに、データに対する DOI^[8]を発行しており、研究者はこの DOI を使用してデータの所在を正確に引用できます。これにより、論文発表後迅速にデータの再利用や参照が容易になり、研究の透明性と信頼性が向上します。

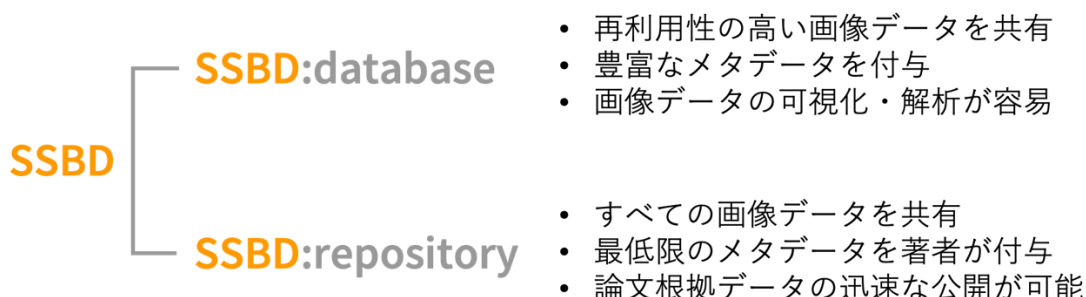


図 2 画像データのエコシステムの構築

研究チームは、SSBD を再構成し、SSBD:database と SSBD:repository の 2 階層のデータリソースを提供することにより、データ公開と共有の柔軟性を向上させた。

一方、2 階層目の SSBD:database は、データの詳細な説明や標準化されたメタデータを含み、他の研究者がデータを容易に理解し、再利用できる環境を整えた公共高付加価値データベースです（図 3）。再利用率の高いデータとして、最先端のイメージング技術で撮影された画像データや、体系的な実験を通じて撮影された画像データを含むデータを収集しています。SSBD:database では、高い専門性を持ったキュレーター^[9]が、データに対する詳細な説明や各種オントロジー^[10]を利用して標準化されたメタデータを豊富に付与します。これにより、データが統一された形式で整理され、異なる研究者や分野を超えたデータの再利用が飛躍的に容易になります。研究者は、豊富なメタデータと標準化されたフォーマットを通じて、必要なデータを効率的に検索・理解できるようになり、新たな知見を得るための時間や労力を大幅に削減できます。また、オントロジーを利用して注釈付けされたデータは、異分野間のコラボレーションを促進し、より広範な問題解決やイノベーションを生むための基盤となります。これにより、科学の進展が加速し、社会全体に対する影響も大きくなるでしょう。

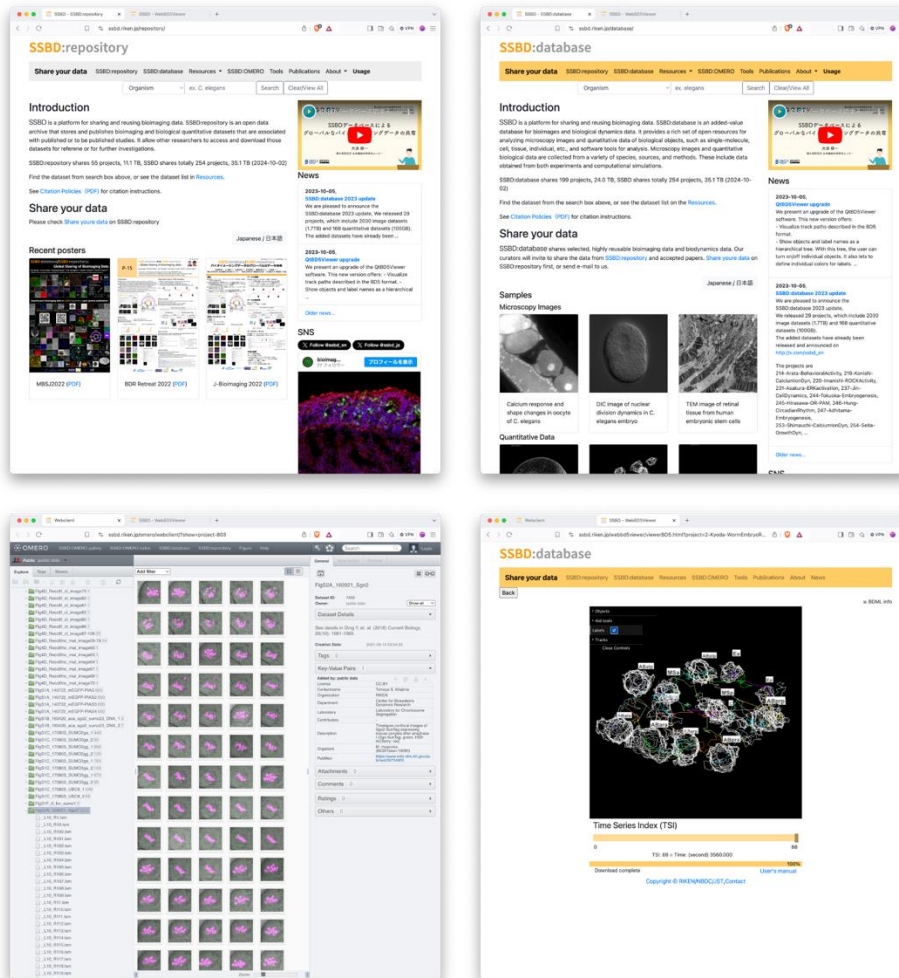


図 3 SSBD:repository および SSBD:database のスクリーンショット

SSBD:repository (左上) と SSBD:database (右上) のトップページのスクリーンショット。SSBD:database では、画像データと定量データの可視化をサポートしており、OMERO プラットフォーム（顕微鏡画像を安全に管理しインターネット上で共有・利用できるフリーのソフトウェアプラットフォーム）を介した画像データの観察（左下）と、独自開発したビューアーでの定量データの観察（右下）をブラウザ上で行うことが可能。

以上のように、従来の SSBD を SSBD:repository と SSBD:database に再構成することで、論文発表後すぐにデータ公開を実現する一方で、データの再利用を促進しやすい構造化されたデータ共有を可能にしました。科学の発展に貢献する生物画像データの共有は世界的な課題となっていますが、SSBD は、日本国内はもちろん、グローバルな画像データ共有のエコシステムの構築において、中核を担うリポジトリおよびデータベースに位置付けられています。現在、日欧豪の七つの組織が参画し、グローバルな画像データ共有のエコシステムを構築するためのコンソーシアム、foundingGIDE が発足しています。研究チームが開発している SSBD と、欧州の画像データのリポジトリおよびデータベースである BiImage Archive^{注1)} および Image Data Resource^{注2)} とのメタデータの調和を図ることにより、各データリソースで共有されている画像データの相互利用を

可能にすることを目指しています。

- 注1) BiImage Archive
<https://www.ebi.ac.uk/bioimage-archive/>
- 注2) Image Data Resource
<https://idr.openmicroscopy.org>

今後の期待

今後、日本国内で撮影される全ての生命科学・医学研究画像データが SSBD:repository で公開されることが期待されます。2024 年に政府の統合イノベーション戦略推進会議で決定された「学术论文等の即時オープンアクセスの実現に向けた基本方針」により、2025 年度以降、競争的研究費の助成を受けた論文とデータの即時公開が義務化されるため、SSBD:repository の重要性は一層高まるでしょう。さらに、SSBD:repository で公開されたデータから、高い専門性を持つキュレーターが、大規模言語モデルなどの技術を活用し、膨大なデータの中から再利用性の高い画像データを選別して、豊富なメタデータと共に SSBD:database で共有する仕組みの構築を目指しています。

画像データの共有により、さまざまな恩恵と派生効果が期待されます(図4)。まず、AI 技術などを利用した画像解析技術のさらなる向上です。公開されたデータを用いることで、多くの研究者が新しい解析手法を開発・検証し、その精度や効率が向上することが考えられます。また、産業界と学術界の両方において、こうした画像解析技術やデータの活用が進むことにより、新しい製品やサービスの開発が促進され、産学連携を通じたイノベーションが生まれることが期待されます。さらに、データがオープンに共有されることで、オープンサイエンスの理念が進展し、科学の透明性と信頼性が高まるだけでなく、広く社会全体にわたる科学の発展に貢献することが期待されます。



図4 SSBD の活用による恩恵と派生効果

SSBD による画像データの共有により、データ再利用・再解析による新たな知識発見、AI 技術などを用いた画像解析ツールの開発、共同研究への発展、人材育成、アウトリーチなど、さまざまな恩恵と派生効果をもたらす。

論文情報

<タイトル>

SSBD: an ecosystem for enhanced sharing and reuse of bioimaging data

<著者名>

Koji Kyoda, Hiroya Itoga, Yuki Yamagata, Emi Fujisawa, Fangfang Wang, Miguel Miranda-Miranda, Haruna Yamamoto, Yasue Nakano, Yukako Tohsato, Shuichi Onami

<雑誌>

Nucleic Acids Research

<DOI>

10.1093/nar/gkae860

補足説明

[1] リポジトリ

データを共有するために整備されたデータベースのこと。オープンな形で一般に共有されたさまざまなデータは、科学分野におけるデータ解析・処理ツールの開発に重要な役割を果たしてきた。近年では、科学研究コミュニティ全体の研究開発を促進するために、研究成果を論文として発表する際に、研究で使用したデータをリポジトリに登録し、広く一般に共有する動きが進んでいる。

[2] foundingGIDE

Euro-BioImaging をコーディネーターとして日欧豪の七つの研究機関・大学が参加する国際的なコンソーシアムにより、Global Image Data Ecosystem (GIDE) の創設を目指す国際プロジェクト。EU の研究・イノベーション支援プログラムである Horizon Europe に採択された。

[3] メタデータ

データを説明するためのデータのこと。大量で形式の異なるデータの横断的な利活用や動画や画像のようなキーワードで検索できないデータに付与されることが多い。

[4] SSBD

研究チームが 2013 年より運用する、生命科学分野の画像データと画像解析により計測される定量データを共有、再利用するためのデータベース。SSBD は Systems Science of Biological Dynamics database の略。

[5] FAIR 原則

FAIR とは「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」の略であり、データの公開と共有を実現するための原則として知られる。

[6] セグメンテーションデータ

画像をピクセル単位で複数の領域に分割し、それぞれの領域に特定のラベルを付けたデータのこと。このデータは、細胞や細胞核などの対象物の輪郭や範囲を明確にする

ために使用される。

[7] トラッキングデータ

時系列の画像や動画内で分割された細胞や細胞核などの各オブジェクトの位置や動きを追跡・記録したデータのこと。

[8] DOI

インターネット上に公開された論文などのコンテンツを、永続的にアクセス可能にするための識別子。DOI は Digital Object Identifier の略。

[9] キュレーター

データセットを収集・整理し、文献などに記述された専門知識をデータセットに注釈付ける役割を担う専門家のこと。

[10] オントロジー

オントロジーとは元来哲学用語で、存在論を意味する。情報科学では知識やモデルに現れる基盤的概念を体系的に定義したものとされ、近年、統制語彙（ごい）や、知識共有および再利用の方法に用いられる。

研究チーム

理化学研究所

生命機能科学研究センター

チームリーダー

（情報統合本部 生命科学データ共有開発ユニット ユニットリーダー）

技師

技師

人材派遣

テクニカルスタッフⅡ（研究当時）

技師

テクニカルスタッフⅡ

パートタイマー（研究当時）

客員研究員

（立命館大学 情報理工学部 計算生物学研究室 教授）

情報統合本部 生命科学データ共有開発ユニット

研究員

（バイオリソース研究センター 統合情報開発室 研究員）

発生動態研究チーム

大浪修一（オオナミ・シュウイチ）

京田耕司（キョウダ・コウジ）

糸賀裕弥（イトガ・ヒロヤ）

藤澤絵美（フジサワ・エミ）

ワン・ファンファン（Wang Fangfang）

ミランダ-ミランダ・ミゲル

（Miranda-Miranda Miguel）

山本春菜（ヤマモト・ハルナ）

中野康恵（ナカノ・ヤスエ）

遠里由佳子（トオサト・ユカコ）

山縣友紀（ヤマガタ・ユキ）

研究支援

本研究は、理化学研究所運営費交付金（生命機能科学研究）で実施し、科学技術振興機構（JST）ライフサイエンスデータベース統合推進事業（統合化推進プログラム）「バイオイメージングデータのグローバルなデータ共有システムの構築（研究代表者：大浪修一）」、同戦略的創造研究推進事業 CREST「データ駆動型解析による多細胞生物の発

生メカニズムの解明（研究代表者：大浪修一）」「体表多様性を創発する上皮-間充織相互作用の動的制御機構の解明（研究代表者：藤原裕展、主たる共同研究者：大浪修一）」、日本学術振興会（JSPS）科学研究費助成事業新学術領域研究（研究領域提案型）「シグニユラリティ細胞の同定と解析のためのインフォマティクス技術の開発（研究代表者：大浪修一）」、同学術変革領域研究（学術研究支援基盤形成）「先端バイオイメージング支援プラットフォーム（研究分担者：大浪修一）」、欧州連合（EU）「HORIZON Europe」研究助成プログラム「Founding a Global Image Data Ecosystem（FoundingGIDE）（研究責任者（日本）：大浪修一）」による助成を受けて行われました。

発表者・機関窓口

<発表者>

理化学研究所

生命機能科学研究センター 発生動態研究チーム

チームリーダー 大浪修一（オオナミ・シュウイチ）

（情報統合本部 生命科学データ共有開発ユニット ユニットリーダー）

技師 京田耕司（キョウダ・コウジ）

技師 糸賀裕弥（イトガ・ヒロヤ）

情報統合本部 生命科学データ共有開発ユニット

研究員 山縣友紀（ヤマガタ・ユキ）

（バイオリソース研究センター 統合情報開発室 研究員）



SSBD チームメンバー

（左から）ミランダ-ミランダ ミゲル、京田耕司、大浪修一、糸賀裕弥、菅原 皓、藤澤絵美



<生命機能科学研究センターに関する問い合わせ>

理化学研究所 生命機能科学研究センター センター長室

報道担当

山岸 敦 (ヤマギシ・アツシ)

<JST 事業に関する問い合わせ>

科学技術振興機構 情報基盤事業部 N B D C 事業推進室

川口貴史 (カワグチ・タカフミ)

Tel: 03-5214-8491

Email: info [at] biosciencedbc.jp

<機関窓口>

理化学研究所 広報室 報道担当

Tel: 050-3495-0247

Email: ex-press [at] ml.riken.jp

科学技術振興機構 広報課

Tel: 03-5214-8404

Email: jstkoho [at] jst.go.jp

※上記の[at]は@に置き換えてください。
