

生物多様性情報の利用

～世界の潮流と日本の現状～

国立科学博物館 植物研究部 細矢 剛

1. はじめに

GBIF（地球規模生物多様性情報機構 Global Biodiversity Information Facility）は、2001年に設置された、インターネットを介して、世界の生物多様性情報を誰でも自由に見られる仕組みを作っている国際機構である。国あるいは国際機関が参加単位となっており、日本も設立当初からの参加者である。演者は、2012年以来、日本ノードのノードマネージャーとして同機構の活動に参加しており、近年のGBIFの総会と生物多様性情報関係の国際的な集會に出席する機会を得て、海外の関連団体等の活動状況などについて情報を得たので、本邦の状況などと比較して紹介する。

2. 生物多様性情報流通のインフラとしてのGBIF

GBIFは、コペンハーゲンに事務局をおく国際組織であり、「生物多様性情報が、科学・社会・継続維持できる未来のために無償で普遍的に利用できる」ことをビジョン、「生物多様性情報の全地球的なリソースとして最重要であり、人類と環境のため、賢く問題解決するためのしくみ」をミッションとして活動を展開している。58参加国・38機関から提供される4万5千件を超えるデータセットに基づき世界中を網羅（陸上ばかりでなく、海上も含めて）しており、現在利用可能なデータは13億を超える。

GBIFが優れているのは、これらのデータの使用を追跡できるテクノロジーである。データがユーザーによってダウンロードされると、そのデータセットにDOI（Digital Objective Index デジタルオブジェクト識別子；コンテンツの電子データに付与される国際的な識別子。コンテンツの所在情報（URI）にも変換され、これでコンテンツにリンクを設定して利用できる）。このDOIを文献と同様に論文中で引用することによって、その論文におけるデータの使用が確実にトレースできる。一方、論文には論文自体のDOIが付与される（例外もある）ので、データセットが論文に紐付けされる。さらに著者にはORCIDのような固有の識別子が、個々のデータにはUUID（Universally Unique Identifier 汎用一意識別子；2つ以上のアイテムが全世界で同じ値を持たないように設定した一意な識別子。自由に生成することができ、生成時刻、PCの固有値、乱数などを利用し、他のUUIDと重複することは起きないようにしている）が与えられることによって、最終的にはIDによって個々のデータがどの著者のどの論文に使用された

かまでが追跡できる。このように、GBIF が提供する生物多様性情報の流通に関するインフラは、データの利活用に重要なものである。

3. 日本の貢献

GBIF にデータを提供するには、IPT (Integrated Publication Toolkit) というサーバーを設置し、GBIF 事務局とを結ぶことが必要である。現在、このサーバーは、国内には国立科学博物館と国立遺伝学研究所に置かれており、この2基のサーバーから750万件のデータがGBIFにデータが提供されている(国立科学博物館は主に標本情報、遺伝研は主に観察情報を担当)。この件数は、アジアでもトップクラスである。しかしながら、GBIF 全体のデータではアジアからのデータ提供は全体の2%にとどまっており、生物多様性が豊かなアジアからは、より一層のデータの提供が求められている。

国内では、科博・遺伝研に加え、東京大学がデータの収集を行っており、この3者の協力によって、国内データが収集されている。現在GBIFに提供されているデータの8割は標本情報である。標本情報は英文ばかりでなく、日本語の情報を伴っているが、GBIFには英語の部分が提供されている。そこで、日本語の部分を利用できるように別途設置されたのが、科博が運営するサイトであるサイエンスミュージアムネット(S-Net)である。S-Netからは101の自然史系博物館、国公立の研究機関、大学から収集された378データセット、536万件のデータの標本情報が公開されており、各データセットや機関に関する情報、学芸員に関する情報なども公開されている。

4. 「ためる」から「使う」時代へ

多数のデータを基にしたデータ解析によって、一点あるいは少数ではなしえなかったような傾向の予測や発見がもたらされる。GBIFはこのようなGBIFデータを使用した科学論文をレビューしており、代表的な使用例が公開されている。これらの中には、多数のきのこ標本の採集情報の解析によって地球温暖化の傾向を示した研究や、アレルギーの原因になる植物の分布予測によって、将来の医療費増大を警告した例、病気のベクターとなる生物と病原体の間の相互関係をネットワーク解析し、防疫上重要な生物について考察した例、など様々な分野(侵略的外来種・気候変動の影響・種の保全と保護地区・開発による影響・生物多様性と人類の健康・食料・農業・バイオ燃料・生物多様性科学の発展・データペーパー・GBIFについての議論など)にわたる文献が多数が紹介されている(最新版は <https://www.gbif.org/ja/document/5Lja8XKRwQDwbhxddOWjtm/gbif-science-review-2019> から公開)。

これに対して、S-Netでのデータの使用例は限られており(H29-30調査)、調査できたのは

23 件にすぎない(<http://science-net.kahaku.go.jp/app/page/activity.html> から公開している)。これらは大きく研究・展示・普及に関わるものに分類でき、研究では分布調査、保全研究における基礎データの調査、分類学的研究における標本の探索 (17 件)、展示に関しては、展示用の標本の探索、資料の作成 (4 件)、普及に関しては、出版用データの作成 (2 件) への利用が挙げられる。使用例の収集が限られているのは、GBIF と異なり S-Net にはデータ使用の追跡機構がないことや、データユーザーには報告義務がないことなども一因であろう。使用の枠をどう広げ、利活用をどのように推進するか、使用例をどう収集するかが今後の課題である。

5. つながる・ひろがる生物多様性情報

現在世界中に様々な生物多様性のイニシアチブが立ち上げられている。これほどイニシアチブが多くなってくると、当然活動の内容に重複が多くなり、コストの無駄が増えてくる。また情報を共有することによって互いに得るものが多くなる場合も多い。そこで GBIF は連携に向けて新しい道を踏み出した。

2018 年 7 月にデンマークのコペンハーゲン大学 バイオテクノロジーセンターで開催された GBIC2 (第二回 Global Biodiversity Informatic Conference) では、世界各国から約 90 名 (日本からは大澤剛士 首都大学・東京准教授、細矢の 2 名) の生物多様性情報学の研究者が集まり、これらの連携について議論された。その結果、連携を推進するための 23 項目の合意が採択され閉会した (表 1; <https://www.biodiversityinformatics.org/ja/call-to-action/> も参照のこと)。GBIF は現在、学名に関する COL (Catalogue of Life)、種情報に関する EOL (Encyclopedia of Life)、文献に関する BHL (Biodiversity Heritage Library)、遺伝子配列に基づいて種を認識する UNITE や BOLD などのイニシアチブと連携を進めており、生物多様性情報のハブとして、さらに機能を強化している。蓄積された情報は、GBIF の中期戦略 (2017-2022) とも合致したものになっており、生物多様性情報のモデル化と利用の基盤となる GEO (地球観測に関する政府間会合) や保全に関する科学データにもとづいた評価を行う IPBES (生物多様性及び生態系サービスに関する政府間科学 - 政策プラットフォーム)、それを実際の条約に落とし込む CBD (生物多様性条約) などを経て、持続的発展を指向した SDG (持続可能な開発目標) の基盤的情報の提供元となっている。

一方、日本では S-Net はジャパンサーチ (<https://jpsearch.go.jp/> 現在は β 版) との連携を進めている。ジャパンサーチは日本国内の連携機関からのデータの提供をうけ、日本の様々なコンテンツを横断検索できるサイトである。現在、約 1960 万件のデータが集積されているが、S-Net に由来するデータは、書籍 (900 万件) について、2 番めの規模である。また、ここに収められたコンテンツは書籍・文化財・メディアアートが中心であり、いわゆる理系資料としては、現在 S-Net が唯一のものである。

ジャパンサーチで生物名で検索を行うと、それを含む自然史情報ばかりでなく、書籍、公文書、

美術、文化財、放送番組、映画などのジャンルのコンテンツがヒットする。博物館資源を今までとまったく異なる視点で利用するためのツールとして、さらに活用されることが望まれる。

6. 今後の発展を目指して

今後の標本・生物多用せ情報学発展を考えると、次のようなことが重要と考えられる。

1) データ範囲の拡大・補間

S-Net は昨年度大幅改修され、インターフェースも大幅に改善された。また、化石データも公開できるようになった。これにより、未掲載の分類群・地域・時間をさらに収集し、集積することが重要である。

2) データの充実

現在、S-Net に画像データを収容することはできない。この背景には技術的な問題だけでなく、法的・制度的な問題もあろう。しかし、画像データへのリンクは設定することができる。画像を入れることは、魅力的なデータベースを構築し、利活用を促進する上でも重要であろう。また、地理情報（緯度経度など）の充実も求められる。

3) デジタルアーカイブへの発展

ヨーロッパには DiSSCO (Distributed System of Scientific Collections) というイニシアチブがある。これはヨーロッパ中の自然史博物館の標本データを統合的に検索可能とする活動で、日本の S-Net と同様の活動といえる。いずれも標本が死蔵されることなく、利活用されることを推進するものである。博物館にとって、標本は最大かつ最重要の資源である。しかし、標本に伴っているのは、採集情報ばかりでなく、それに由来する利用履歴（論文・貸し出し）や展示履歴など様々なものである。一般に、デジタル化した情報は、互いの親和性が高くなる。標本に関連した情報がデジタル化されていれば、標本情報と関連付けて、利活用の幅が広がる。博物館の使命として、標本資源をさらに利活用し、いわゆる「知的循環」を超える「知的拡大再生産」に活かすことが今後求められる。

表1 GBIC2 において提案された生物多様性情報の連携に向けての提言 (大澤ら (2019) を改変)

科学と証拠に基づく計画立案の支援

1. 生物多様性の知識および知見を提供する上で、研究において重要な要件を満たした形式であり、かつ社会の目指すゴールに対して生物多様性が正しく測定・評価できる形式であること。
2. 自然の体系の機能や状態に関する知見が得られる生物多様性情報科学の基礎研究の基盤を提供すること。
3. 現有の知識を保存し、それに基づいて改善することによって、生物多様性の知見をさらに広げるプラットフォームを提供すること。

オープンデータとオープンサイエンスのサポート

4. 生物多様性データの公開において、フリーでオープンなデータ共有に対する障壁を取り除き、FAIR 原則 (Wilkinson et al. 2016: <https://doi.org/10.1038/sdata.2016.18>) にしたがうこと。
5. すべてのデータリソースは、現在および将来の再利用のため、豊富なメタデータを伴って記述すること。
6. すべてのデータリソースは、安定して、永続的で、信頼できるリポジトリに必ず保管されるようにすること。
7. 関係した専門家や専門家のコミュニティが協同作業によるデータのキュレーション、アノテーションおよび改善を行えるようにすること。
8. 専門的知識の提供に貢献したすべての人に対して、その貢献が完全に記録され、謝辞に述べられ、クレジットされるようにすること。
9. すべての情報のソースの出所と帰属が追跡可能であること。

高度に連結する生物多様性データのサポート

10. 博物館コレクションや文献など、歴史的なデータソースの構造化されたデジタル情報を利用できるように集約すること。
11. 新規の観察・測定情報は、取得後可能な限り早く、構造化されたデジタル情報としてアクセスできるようにすること。
12. 異なるクラスの生物多様性情報 (分布・形質・遺伝子など) を全体に相互連結した形で組み合わせたり、検索したり、解析したりできるようにすること。
13. 他の研究コミュニティやインフラとの連携により、地球観測、社会科学的データ、その他のリソースとの相互運用性を実現すること。

国際的なコラボレーションのサポート

14. すべての地域とすべてのセクターで、生物多様性情報学に関する能力強化のニーズを解決すること。
15. 提供されている知識インフラのうち、コミュニティが不可欠な要素として認識するサービスおよびコンポーネントを維持するための資金を確保すること。
16. 前述の知識インフラを構成するすべてのコンポーネントをデザインし、構築し、維持する上で、柔軟かつ協働可能なアプローチを考案すること。
17. インフラ・ツール・サービスの進歩、事例の積み重ね、能力の向上がすべての国・地域の関係者に行き渡り、その利益を享受できること。
18. データをつくる段階から解析や応用に至る全工程で、すべての地域のすべての関係者グループが参加でき、協力し合えるようにすること。
19. すべての国・地域において科学および政策決定の土台となるデータのレパトリエーション (原産国にデータを受け渡すこと) が可能であること。
20. すべてのスケール (世界、地域、国、地方) でデータへの効果的なアクセスと利用が保証されること。
21. 国際的なソリューションにおいて不可欠で効果的な要素である地域・国・地方レベルでの投資に対して謝意と支持を表明すること。
22. 言語・文化の壁を超えてデータが共有や使用がなされること。
23. ABS (access and benefit sharing) に鑑み、国際的な合意の実行をサポートすること。

