

GBIF 経由で公開される種発現 (occurrence) データセットの使用許諾

2014 年、「GBIF 運営委員会」は、GBIF.org 経由で共有されるデータに関し、データ発行者とデータ使用者の双方の使用方法をより明確にする必要性に取り組みました。2013 および 2014 年度の「事務局」のコミュニティ協議 ([/newsroom/consultations#licensing](#)) を受け、「運営委員会」は次の一般方針を策定しました。

ネットワーク内の全ての種の発現データセットを、「Creative Commons」による以下の 3 つの選択肢のうちの 1 つに相当するデジタルライセンスに確実に関連させること。

- ユーザー側に対する制約や特定要件を伴わず、データを使用可能とする際に準拠する「CC0」
- データを使用可能とする際に準拠する「CC-BY」。ただし、使用データの出典には所有者指定の方法で帰属が適切に付されることが条件
- データを使用可能とする際に準拠する「CC-BY-NC」。ただし、帰属が適切に付され、かつ、営利目的に使用しないことが条件

「GBIF」は、データ発行者が利用可能な最もオープンな選択肢（「CC0」）を選択することを強く推奨します。「GBIF」コミュニティのメンバーである Peter Desmet は 2012 年にその根拠を簡潔に説明しています

(<http://www.canadensys.net/2012/whyweshouldpublishourdataundercc0>) 。

まず、当組織のデータ、データセットおよびデータベースにて著作権登録の対象はほとんどありません。著作権はクリエイティブなコンテンツにのみ適用されるもので、当組織データの 99% が事実ベースであるため、著作権保護は不可能です。当組織は摘要欄のテキスト、選択／作成したデータ・フォーマットまたはデータベースモデル、ならびに画像に係る著作権を保有しています。（当組織がデータを公開する方法において準拠する）「Darwin Core Archive（ダーウィン・コア・アーカイブ）」を考慮すると、クリエイティブなコンテンツの量はさらに少なくなります。すなわち、データ・フォーマットは規格であって当組織が提供するの画像そのものではなく、そのリンクに過ぎません。

コンテンツの所有者にとって、事実が終了する時点と（著作権保護が可能な）クリエイティブなコンテンツが開始する時点とを解明するのはもはや困難かもしれません。そこでユーザーにどのような法律上の悪夢が起りうるか想像してみましょう。また、国によって適用される規則は異なります。「CC0」に基づいたデータ公開は曖昧さや形式主義というものを取り除いてくれます。当組織はクリエイティブなコンテンツに係る著作権保有を放棄し、当組織のデータはパブリック・ドメインという法的立場をとっています。もはや誰も著作権保護はできません。

.....クレジットを付するのは効果的でしょう!.....ただし、これを法的に強制することで正反対の

影響が生じる可能性があります。つまり、ユーザーが使用許諾を完全に順守していないのではなく、いかという不安からデータを使用しなくなることが考えられます（上段を参照してください）。本記事の冒頭で指摘しましたように、「CC0」は帰属を付するという過度な法的強制要件をなくしますが、帰属を付するという道義的責任を軽視するものではありません。実際、これは今まで数十年以上にわたって科学研究分野における慣習となってきました。すなわち、法的には使用する研究／データを引証する必要はありませんが、かといって、引証しないことで盗用であると考えられる可能性があり、結果的に、研究の評判や信頼に傷が付くことになります。

.....[CC-BY-NC に準拠すると]、ユーザーは営利目的でデータ（セット）を使用することができません。これは学問的見地からは望ましいように思われますが、使用許諾というのは直感的思考とは比較にならないほど数々の制約を伴います。Hagedorn, G. et al. 2011 (<http://zookeys.pensoft.net/articles.php?id=3036>) を参照してください。

商業的ジャーナルを出版するためにデータを再販売するなど明確な営利目的での使用から、費用回収の手段として広告を掲載するウェブサイトといった作品や成果物まで、何を基準に営利／非営利目的の使用と考えるのかについては、著しく解釈が相違しています。

標準化デジタル使用許諾の有用性

「GBIF」の使命は生物多様性データへの無償のオープン・アクセスを推進することです。公開当初にデータ制約が導入されると、大きな問題に対処するためデータセットが統合されるにつれ、制約が増えていきます。

大規模な研究で重要なのは、一般的に、できるだけ制約のない「ビッグ・データ」の基盤です。例えば、Global Biodiversity Informatics Outlook (<http://www.biodiversityinformatics.org>) は生物多様性情報科学の重要な基盤の1つとして「オープン・アクセスと再使用の文化」を挙げています。Research Data Alliance (<https://rdalliance.org>) 等のグローバルなイニシアチブは、研究データ全般において同様の課題に取り組んでいます。これまで以上にデータ（特に公的資金注入によるデータ）へのオープン・アクセスが可能になるよう、方針の策定を進めている国も数多くあります。このようなミッションのもと、「GBIF」はこういった傾向をより良く理解する立場にあり、責任を担っているのです。

これまで、「GBIF」データのユーザーは「データ使用契約」 (</terms/licences/datause>) に従うことに同意し、同様に、データ発行者は「データ共有契約」 (</terms/licences/datasharing>) の条項を順守することに同意してきました。こうした双方の契約は、データ発行者がデータセットに追加しかつメタデータに盛り込む付加的な「規約条件」に言及しています。

その結果、「GBIF」の多くのデータ発行者は各データセットに係るメタデータに自身の使用許諾の詳細を記載しています。こうした付加的な規約条件はフリーテキスト形式で提出され、多様な非標準フォーマットで書かれているため、「GBIF.org」のユーザーにとって、大容量データのダウンロード、特に複数の機関からのデータを伴うダウンロードに係る諸権利を把握することが不可能とは言えないまでも難しくなっています。こうした条件が大きな問題となるのは、自動ツールのウェブサービス経由でデータが使用される際、使用上の制約が人的に読み取り可能なテキスト形式のみでまとめられおり、ツールがそれを構文解析できない場合です。

データのユーザーにとっての意味

データのユーザーにとって、使用許諾一式が標準化されかつ制約化されるということはすなわち、データを安心して使用できるようになるということで、データの（再）使用が促されます。ユーザーはそもそも、1つ1つ使用許諾に目を通してデータ使用が許可されるかどうかを判断しなければならず、不明瞭な場合にはデータ使用をやめようと考えます。使用許諾一式の標準化および制約化は、このようなユーザーの負担を取り除くことになるのです。「GBIF」のような情報ネットワークが適切なフィルタリングを行うことで、ユーザーは、ダウンロードしたデータセットが使用許諾にある用途での使用が可能かどうか、容易に分類し限定できるのです。ユーザーが認識しておかなければいけないのは、「CC0」に言及するデータセットであっても、「GBIF」のデータ使用契約 (<http://www.gbif.org/terms/licences/datause>) が依然として適用されることです。一例を挙げると、出典の適切な帰属と引証が必要です。これは新たな引証メカニズムによってサポートされます。出版物等への気軽な引用では、ダウンロード物に含まれるデータセットに一貫した参照が供与されます。

データ発行者にとっての意味

「GBIF」は、発行者が種の発現データの収集、デジタル化、キュレーションおよび公開にあたってデータ使用を実証しなければならず、付随する取り組みや努力に対してしかるべき評価を受けたがっていることを認識しています。

より透明性が高く、かつ制約の少ない使用許諾アプローチの導入を補完するために、データ使用の追跡（トラッキング）と報告をサポートし、データ使用者が容易に引証を行えるツールを提供するため、必要なあらゆる取り組みが行われます。使用許諾の順守を強制しようとするのは問題ですが（上記の適用性の検証、ならびに限定的な法的効力を参照してください）、規範を策定し、コミュニティの最良実践を支援することで、帰属の必要性が理解され、最良実践が適切なツールと手順に支えられていくことになります。

「GBIF」は「デジタル・オブジェクト識別子（DOI）」システムを基盤とする引証モデル開発にあたり、作業計画に取り組んでいます。これが導入されると、各データセットには「DOI」が付され、引証は全て永続的な「DOI」を通じてデータセットを参照することになります。こうした引証モデル下では、データのユーザーに対し、ダウンロード対象に係る引証の容易な、かつ安定した出典参照（関連データセット全てへの参照を含む）を供与することが可能になります。出版印刷物やデジタルアプリケーションで使用されたデータを追跡する発行者の能力を向上するモデルです。

他の使用許諾

使用許諾の中には、上述した3つにほぼ相応すると考えられるものもあります。（「CC0」に非常に似た）「PDDL」と（「CC BY」に相応する）「ODC-By」が主な例です。こうした使用許諾下で公開されるデータセットを推奨するには、対応する同等の「CC」使用許諾を考慮すべきです。先に引用した論文

（<http://www.canadensys.net/2012/whyweshouldpublishourdataundercc0>）では、他のオープンデータ使用許諾の長所と短所を考察し、比較検討しています。ただし、「GBIF」経由で公開されるデータは上述の3つの「CC」バージョンの1つに分類される必要があります。

GBIF の文脈での「営利使用」の意味

営利使用の定義については多様な解釈があります。例えば、商業用ジャーナルの出版物での使用と比較して、対象の使用行為が、データの再販売といった営利行為に該当するかどうかで絞り込むという方法もあります。また、広義の解釈では、運営管理上の費用回収の手段として広告を掲載するウェブサイトにもまで範囲が広がられます。当組織にはこういった議論の解決策を提案したり強要したりする意図はありませんが、本項では、「営利」と「非営利」の指定に係る当組織の

現段階での実践について明記します。

「GBIF」は、科学的理解の支援および生物多様性と保護保全の向上を目的としたオープン・アクセス型インフラを提供しています。当組織は「GBIF」が介在するデータを引用する標準的行為としての帰属（「CC-BY」）の普及に今後も取り組みますが、これは原書の引証にあたり、当組織が対象とするコミュニティにおいて認知された規範を反映していると確信しています。新規版の「統合パブリッシング・ツールキット（IPT）」は、機械読み取り可能な標準的使用許諾として最も制約的で、発行者はデフォルトにて「CC-BY-NC 4.0」を選択できます。ユーザーも発行者も、営利使用と非営利使用の定義におけるグレー領域を認識すべきです。当組織は、非営利使用の制限的な解釈が理念、ならびにオープン・アクセス全般およびとりわけ「GBIF」の意義に反することになると確信しています。

Page4

ユーザーは善意を持って「CC-BY-NC」使用許諾を伴うデータを使用すべきです。当組織がユーザーに強く推奨するのは、非営利使用に対する広義でありながらも合理的な解釈によって、彼らが受け取る恩恵が事実上の営利となるかどうか、発行者と意思疎通を図ることです。

当組織は発行者による「CC-BY-NC」使用を支援し、そうした使用許諾の意図を重視すべく、合理的な取り組みを実践します。より自由に使用許諾される集約データなどにまで制約が加えられるようなケースでも同様です。

ただし、「GBIF」は法的な手段で「CC-BY-NC」を強要する利害もリソースも持っていません。ユーザーが使用許諾に対して重大な違反をする、または悪意の行為をするような場合、発行者は法的措置を講じることができますが、「GBIF」は関与しません。こうした状況に満足いただけない発行者には、共有するデータ・エレメントを制限（「範囲縮小」）するか、または「GBIF.org」経由でのアクセス対象からデータを除去するか、いずれかを実行していただきたいと思います。