

Title	オープンデータプラットフォームがもたらす触媒効果：グローバル生物多様性情報ファシリティ(GBIF)の事例
Author(s)	沼尻, 保奈美; 林, 隆之
Citation	年次学術大会講演要旨集, 38: 80-84
Issue Date	2023-10-28
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19210
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

オープンデータプラットフォームがもたらす触媒効果：グローバル生物多様性情報ファシリティ (GBIF) の事例

○沼尻 保奈美、林 隆之（政策研究大学院大学）

1. はじめに：オープンデータと触媒効果

オープンサイエンスは、研究の透明性と再現性を確保する手段として、多くの先進国の科学技術政策や学術政策の主要なアジェンダとなっている。2023年5月に開催されたG7 仙台科学技術大臣会合において採択された「G7 科学大臣コミュニケ」（共同声明）ⁱにおいて、主要国によるオープンサイエンスの推進が強調された。また、EUにおいても、Horizon 2020 や次期フレームワークプログラムにおいて、オープンサイエンスの取り組みが基本方針として掲げられている（European Commission, 2018）。その一方で、オープンサイエンスは、研究成果の信頼性を高めるのみでなく、学際的な共同研究の機会を増加させる可能性があると考えられている。しかし、限られた科学技術予算や人的リソースの中で、オープンサイエンスの取り組みが学術界における研究の効率性や加速にどの程度寄与するかについての明確な結論はまだ得られていないⁱⁱ。

オープンデータは、オープンサイエンスの中心となる要素の一つとされ、研究過程で得られたデータの公開によって、研究の透明性の向上やアクセス性の拡大を促進することが期待されている^{iiiiv}。特に、現代の学術研究においてデータ駆動型のアプローチが増加する中、オープンデータはオープンサイエンスの推進とその加速に不可欠な要素として位置付けられている^v。実際に、いくつかの研究機関や資金提供機関では、オープンデータは標準的な慣行と見なされつつある^{vi}。研究者が研究過程で得たデータを公開するオープンデータに関する研究については、これまでFigshare リポジトリのオープンな研究データの利用状況についての研究^{vii}などが行われている。

その一方で、オープンデータには、研究者によるデータ公開のみではなく、国などの公的機関が生成・保有するデータを公開するオープンガバメントデータ (OGD) が存在する。これらのOGDは、新しい学問的探究や研究の遂行を促進するための貴重な情報源としての位置付けがなされている。例えば、アメリカのUSGSは、OSTPが2014年3月に発行した物理的な科学コレクションを保存・管理し、研究資産として活用する指針を元に、国立地質地球物理学データ保存プログラムなどにより地質情報をデータベース化し、OGDとして共有している^{viii}。しかし、政府がデータを公開していても、その公開範囲が制限されている場合もあれば^{ix}、これらのデータが異なる機関やプラットフォームに分散して保存・公開されている現状があり、このOGDの分散性はデータの発見や再利用を困難にするとともに、研究活動の効率性を低下させる要因となっている。OGDを活用した政府の活動の評価についての研究^xや、行政官と市民のOGDの活用事例^{xi}は存在するものの、現状では研究者が具体的にOGDをどのように活用して研究を進めているか、その詳細な実態は十分に明らかにされていない。

上記を解決するために、個々のデータを一箇所に集約するオープンデータプラットフォームが新たな研究開発や研究者の共同を促進する場として注目を集めている^{xii}。

このようなオープンサイエンスの形は、研究助成金のように直接的に研究推進を促すものではないが、研究者の研究活動に影響を与え、通常は独立して交流のない異分野間の研究者同士が共同で研究を行うなど、間接的に研究活動の構造的な変化を促進する仕組みであると期待される。本研究では、このような間接的な影響を「オープンサイエンスの触媒作用：Open Science catalysis」（以下、触媒作用とする）と独自に称する。この「触媒作用」とは、研究者の意識や方法論に変化をもたらし、その結果として異分野間の研究者との連携が促進されるというものである。

特に、地球の環境変動や生態系の保護といった課題に直面する現代において、生物多様性に関するデータはその重要性を増している。生物多様性データは、種の分布、生息地、生態系の健康状態など、地球上の生命の状況を把握するための基盤となる情報を提供する。これにより、絶滅の危機に瀕している

種の保護や、生態系の維持・回復のための政策策定の指針となる^{xiii}。グローバル生物多様性情報ファシリティ (GBIF) は、この文脈において、オープンデータプラットフォームの先駆的な取り組みの一つとして位置付けられる。GBIF は、全世界からの生物多様性データを集約し、公開することで、科学者や政策立案者に情報を提供している。本稿では、オープンデータプラットフォームがもたらす触媒効果を明らかにし、特に GBIF の事例を通じて、その実際の影響と利益を考察する。

2. 分析：オープンデータプラットフォームの触媒効果

2.1 研究の問い

オープンデータプラットフォームが研究者の研究内容に対してどのような影響をもたらすのかが不明な現状を踏まえ、以下の問いを立てる。オープンデータプラットフォームを利用した経験のある研究者において、プラットフォームの利用前と利用後で、出版された論文の研究分野および研究内容にどのような変化が見られるか。

2.2 分析対象：グローバル生物多様性情報ファシリティ (GBIF) とは

本研究で対象とする GBIF とは、生物多様性データを収集・公開する国際的なネットワークおよびデータベースのことを指す¹(Research and societal benefits of the global biodiversity information facility)。GBIF は OECD のメガサイエンスフォーラム生物多様性情報学サブグループによる 1999 年の勧告を元に 2001 年に設立された。GBIF は、膨大な量の生物多様性データを一点に集め、公開することによって、生物調査の重複を避け、効率的に科学研究を行い、経済的および社会的利益を生み出すことを目的としている²。GBIF データベースには、動植物、菌類、微生物などの多岐にわたる生物種の情報が含まれており、これには生物の分布情報、分類学的情報、生態学的情報などが含まれる。データは世界中の博物館、研究所、大学、政府機関などから提供され、データを標準化された形式で提供し、オンラインでアクセス可能な API の提供により、研究者や開発者は GBIF のデータを自らの研究に簡単に組み込むことが可能である。2023 年 2 月

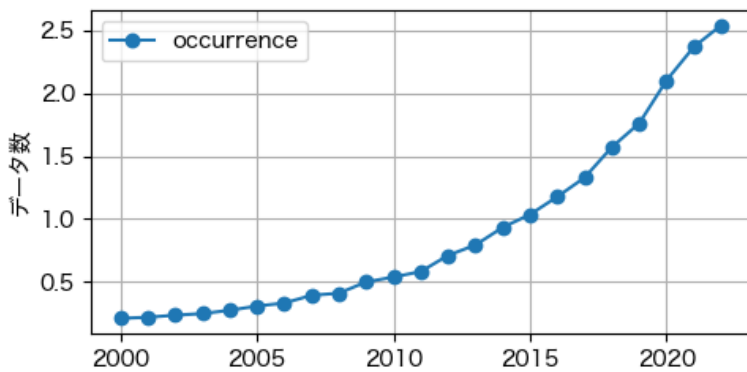


図 1 occurrence データ数の時系列変化

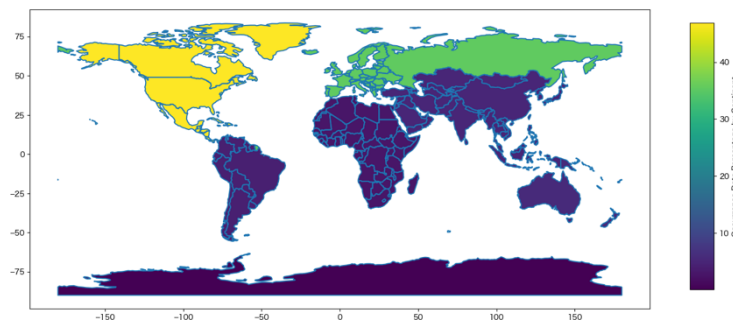


図 2 データの提供元国の割合

時点では、約 23 億の occurrences (生物の出現情報)、species (生物種の分類情報)、datasets (データセットの出典元の情報) などのデータが集積されている^{xiv}。図 1 において、2000 年から 2023 年現在までの occurrences データの収録数を示す。2010 年ごろから収録データは年々増加している。さらに、図 2 において、全世界の収録元国の割合を示す。GBIF に収録されているデータはその半数の 46% が北アメリカからなるデータであり、アフリカは約 2% ほどしかない。

GBIF は、生物多様性に関する研究や政策策定のための基盤として、世界中の研究者や意思決定者に利用されている。特に、種の保全、生態系の健康、気候変動の影響分析などの分野で、GBIF のデータは重要な役割を果たしている。

¹ <https://clarivate.com/ja/solutions/data-citation-index/><https://clarivate.com/ja/solutions/data-citation-index/>

² <https://www.gbif.org/what-is-gbif>

GBIF のサイトには、研究者が GBI の研究を利用して行われた論文のタイトル、DOI および論文のリンクがデータベース化されている。2023 年 5 月時点で、GBIF に掲載されているデータを使用した研究論文の DOI は 6,876 件であった。この DOI から、Scopus を検索し 7,052 件の論文情報を取得した（以下、GBIF 使用論文）。図 3 には、GBIF 使用論文の上位 8 研究分野（大分類）の年次変化を示す。主要な分野はエコロジー複合分野であり、GBIF 使用論文は年々増加している。また、その他の全ての分野において増加傾向であるが、エコロジー複合分野ほどではない。

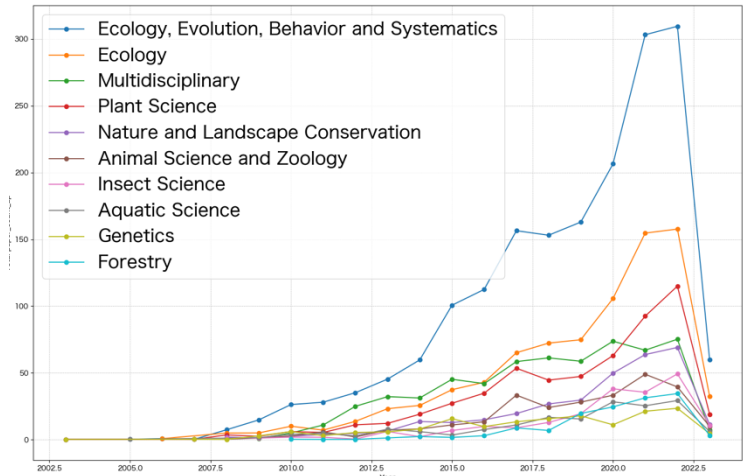


図 2 各分野における GBIF 使用論文の推移

3 分析 1：GBIF を使用した研究者の分野移動

研究者が GBIF を利用した前後で、研究者が出版する論文の分野に変化が生じているかを明らかにするために、一度でも GBIF を使用したことがある研究者がこれまでに出版した全論文を取得し、初めて GBIF を使用した論文の出版前後で分野がどれだけ移動しているかを確認する。

GBIF 使用論文群から研究者 ID を取得し、それらの研究者 (GBIF 研究者) が現在までに出版した研究論文情報 5,480,571 件を取得した (GBIF 研究者論文)。次に、各 GBIF 研究者が出版した全論文を、GBIF を初めて使用した出版年を境に、GBIF 前と後で分ける。

本報告では、GBIF を初めて使用して書いた論文の数が最も多い 2017 年において、それらを出版した研究者 540 人を対象にする。しかし、GBIF 使用前後で分割して分野の移動距離を測定したとしても、長期的な移動トレンドを反映しているだけの可能性もあるため、GBIF を使用する前 5 年およびそれよりもさらに前の 5 年の計 10 年を設定し、分野間移動距離を測定することによって、トレンドの変化があるかを確認することにする。分野間距離の移動を計算する際には、各期間の分野別論文数を取得し、Scopus 収録論文全体の分野分布から内在距離を計算したベクトルで重み付けを行った上で、コサイン距離を算出する。

4 分析 2：GBIF を使用した研究者の研究分野の変化

2007 年～2011 年と 2012～2016 年、2012～2016 年と 2017 年～2021 年の期間における研究者の分野移動距離を図 4 および図 5 に示す。分野 11 (General Agricultural and Biological Sciences)、分野 23 (General Environmental Science)、分野 27 (General Medicine) においては距離に変化はほぼ見られない。一方で、分野 19 (General Earth and Planetary Sciences) と分野 30 (General Pharmacology,

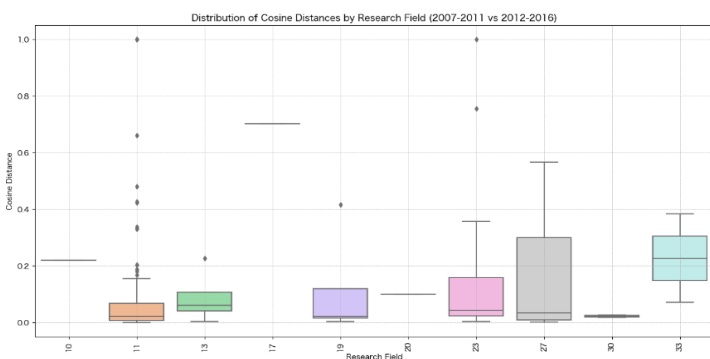


図 4 2007 年～ 2011 年と 2012 年～ 2016 年における研究者の分野移動距離

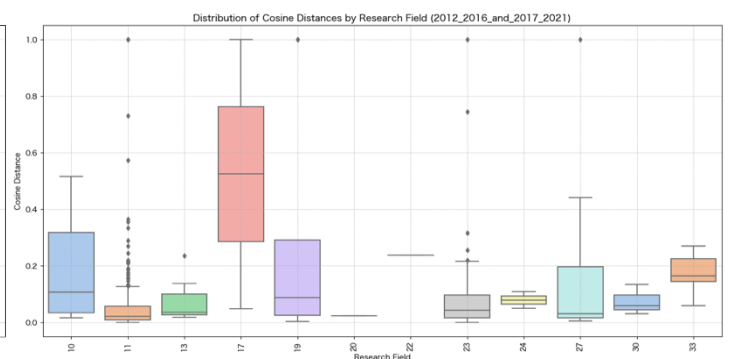


図 5 2007 年～ 2011 年と 2012 年～ 2016 年における研究者の分野移動距離

2017年にGBIFを初めて利用した研究者について、GBIF利用前(2012~2016年)と利用後(2017~2021年)での分野移動距離が0.1より大きい研究者80人の論文1,124件を対象に、研究トピックがいかに変化しているかを確認した。図5はGBIF利用前(2012~2016年)であり、「Forensic Entomology; Forensic Science; Chrysomya Megacephala」(法医学昆虫学)のトピックの割合が大きい(2.16%)。一方でGBIF利用後(2017~2021年)においては、「Maximum Entropy; Ecosystem; Environmental Space」(環境空間)が一番トピックの割合が大きいように変化している(0.71%)。

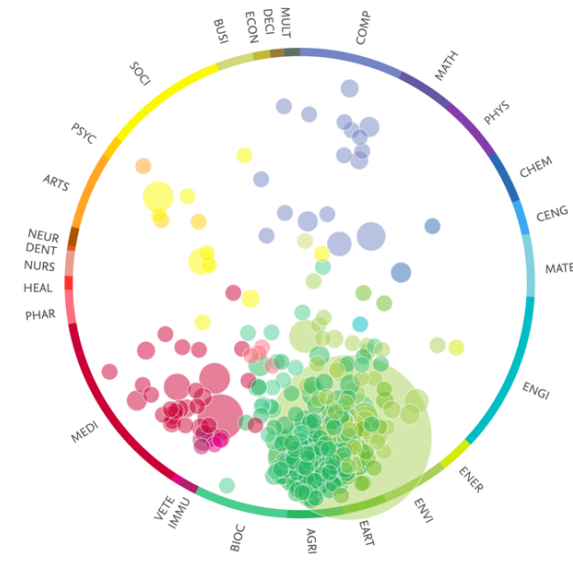
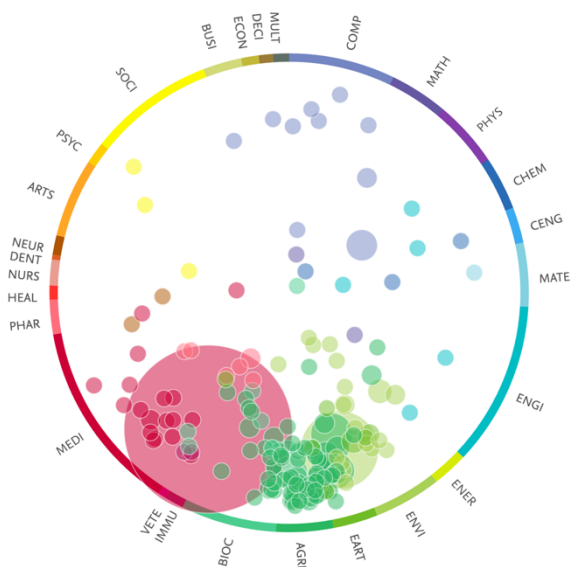


図 6 2012 から 2016 年における SciVal のトピック

図 7 2017 から 2021 年における SciVal のトピック

ただし、このような変化はGBIFの利用とは関係なく、「Forensic Entomology; Forensic Science; Chrysomya Megacephala」の研究が、2017年以降には全体的に研究トピックを変化させてきた可能性も否めない。そのため、対照群との比較を行う。2017年以降にGBIFを使用していない「Forensic Entomology; Forensic Science; Chrysomya Megacephala」トピックの研究を行っている研究者のうち、

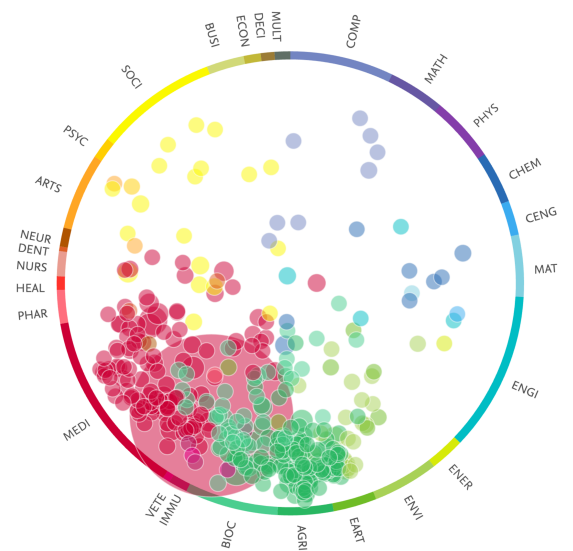
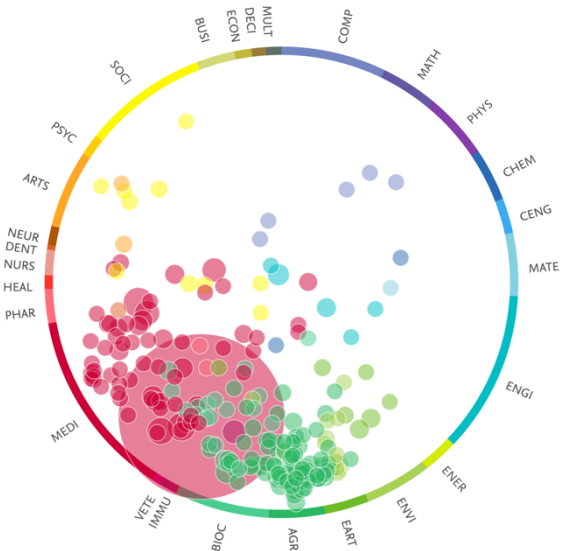


図 8 2012 から 2016 年における対照群の SciVal のトピック

図 9 2012 から 2016 年における対照群の SciVal のトピック

研究者の属性(所属国・所属機関)がGBIF利用研究者の構成と近くなるようにランダムサンプリングを行い、それらの著者がこれまでに出版した論文群のトピックについて、同様に2012~2016年と2017年~

2021年の期間における Scival のトピックの変化を確認した(図 7, 図 8)。結果、GBIF を使用していない研究者については、Forensic Entomology; Forensic Science; Chrysomya Megacephala から大きな変化は確認されなかった。

5. 議論

本研究では、オープンデータプラットフォームによる「触媒作用」が研究者の研究内容に影響を与えるかを分析した。Scopus の 27 分野レベルでは一部を除き大きな変化を確認することはなかった。ただし、比較的移動距離が大きい研究者が有するトピックは、GBIF を利用したことのない研究者と比べ、GBIF を利用した研究者のトピックに変化があることが明らかになった。特に 2012 年から 2016 年の割合が大きかった Forensic Entomology; Forensic Science; Chrysomya Megacephala というトピックは、法医学における昆虫の研究であり、発見される昆虫の種類や発育段階を調査することで、死後の時間経過や死体の移動履歴などの情報を推定するものである。一方で、2017 年～2021 年において、変化した「Maximum Entropy; Ecosystem; Environmental Space」(生態学)というトピックは、生態系の動態のモデル化や環境空間のデザインが生態系に与える影響を評価する研究を指す^{xv}。このような研究では、大量で広範囲のデータが必要とされ、それらのデータは GBIF に収録されている。よって、元々は特定の昆虫に関する研究を行っていた研究者が、GBIF のウェブサイトを訪れ、データを使用したことにより、大量で大規模のデータを必要とする研究を行うようになった可能性がある。

Forensic Entomology; Forensic Science; Chrysomya Megacephala というトピックは特殊なトピックであるため、今後は GBIF を初めて利用した研究年を広げるなどして、その他のトピックについても分析が必要である。さらに、GBIF のみではなくその他の分野のプラットフォームについても分析を進めるなど、オープンデータプラットフォームによる研究者への影響について明らかにしていく。

参考文献

- ⁱ https://www8.cao.go.jp/cstp/kokusaiteki/g7_2023/230513_g7_communique.pdf
- ⁱⁱ Woelfle, M., Olliaro, P., & Todd, M. Open science is a research accelerator. *Nature Chem* **3**, 745-748 (2011). <https://doi.org/10.1038/nchem.1149>
- ⁱⁱⁱ Koznov, D., Andreeva, O., Nikula, U., Maglyas, A., Muromtsev, D., & Radchenko, I. (2016). A Survey of Open Government Data in Russian Federation. , 173-180. <https://doi.org/10.5220/0006049201730180>.
- ^{iv} Kansa, E. (2012). Openness and archaeology's information ecosystem. *World Archaeology*, 44, 498 - 520. <https://doi.org/10.1080/00438243.2012.737575>.
- ^v Ramachandran, R., Bugbee, K., & Murphy, K. (2020). From Open Data to Open Science. *Earth and Space Science*, 8. <https://doi.org/10.1029/2020EA001562>.
- ^{vi} Maeda, E., & Torres, J. (2012). Open Environmental Data in Developing Countries: Who Benefits?. *AMBIO*, 41, 410-412. <https://doi.org/10.1007/s13280-012-0283-4>.
- ^{vii} Quarati, A., & Raffaghelli, J. (2020). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, 48, 423 - 448. <https://doi.org/10.1177/0165551520961048>.
- ^{viii} National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Board on Research Data and Information; Committee on Toward an Open Science Enterprise. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington (DC): National Academies Press (US); 2018 Jul 17. 3, The State of Open Science. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK525413/>
- ^{ix} Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data - A critique of open government datasets in the UK. *Gov. Inf. Q.*, 37.
- ^x Boudreau, C. (2020). Reuse of open data in Quebec: from economic development to government transparency. *International Review of Administrative Sciences*, 87, 855 - 869. <https://doi.org/10.1177/0020852319884628>.
- ^{xi} Ruijter, E., Grimmelikhuijsen, S., Berg, J., & Meijer, A. (2020). Open data work: understanding open data usage from a practice lens. *International Review of Administrative Sciences*, 86, 19 - 3. <https://doi.org/10.1177/0020852317753068>.
- ^{xii} Ojo, A., Porwol, L., Waqar, M., Stasiewicz, A., Osagie, E., Hogan, M., Harney, O., & Zeleti, F. (2016). Realizing the Innovation Potentials from Open Data: Stakeholders' Perspectives on the Desired Affordances of Open Data Environment. , 48-59. https://doi.org/10.1007/978-3-319-45390-3_5.
- ^{xiii} Heberling, J., Miller, J., Noesgaard, D., Weingart, S., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 118. <https://doi.org/10.1073/pnas.2018093118>.
- ^{xiv} https://gbif.jp/about/jbif/summary/pdf/jbif_pamphlet_20230315.pdf
- ^{xv} <https://royalsocietypublishing.org/doi/10.1098/rstb.2010.0018>