



学術情報

「オープンデータ」という考え方と、生物多様性分野への適用に向けた課題

大澤 剛士^{1,3,*}・神保 宇嗣^{2,3,4}・岩崎 亘典^{1,5}

¹ 独立行政法人農業環境技術研究所

² 国立科学博物館

³ GBIF 日本ノード JBIG

⁴ Linked Open Data for Academia (LODAC)

⁵ OSGeo 財団日本支部

Current status and future perspective on “Open Data” in biodiversity science, Japan

Takeshi Osawa^{1,3,*}, Utsugi Jinbo^{2,3,4}, Nobusuke Iwasaki^{1,5}

¹ National Institute for Agro-Environmental Sciences,

² National Museum of Nature and Science,

³ Japan Node of Global Biodiversity Information Facility,

⁴ Linked Open Data for Academia (LODAC),

⁵ The Open Source Geospatial Foundation Japan chapter

キーワード：オープンソース、生物多様性情報、相互運用性、データ公開、データ発掘、ライセンス、GBIF

はじめに

科学において、論文で使ったデータをはじめ、基礎的なデータをデータベース等に集約して公開することは日常的に行われている。それらの情報は、ただ閲覧するだけではなく、条件に従って再利用することもできる。生物分野の著名な例としては、国際的な塩基配列データベースがあげられる (International Nucleotide Sequence Database: コラム 1)。これら塩基配列データベースには、多くの研究者が、自身が決定した塩基配列を登録するとともに、登録されている遺伝子情報を利用して研究を進めている。このような「データ公開」と、その「再利用」は、科学だけでなく、様々な分野で実施されている。

科学に限らない「データ公開」の一般論として、オープンデータ (Open Data) という新しい考え方が急速に広がり、一般に定着しつつある。オープンデータとは、データを「無償かつ自由に利用可能にする」という考え方

*e-mail: arosawa@gmail.com

であり、データをこの考え方に沿って公開することをデータの「オープン化」と呼んでいる (Open Knowledge Foundation 2012 Open Data Handbook Documentation, Release 1.0.0. <http://opendatahandbook.org/> 2014 年 3 月 4 日確認)。データのオープン化は、データの加工と再利用を促進することで、社会や経済の活性化・効率化を図る事、行政データの公開を通じた透明性確保、市民の行政への参加促進とそれに伴う行政サービス向上等が主要な目的とされている。昨今では、オープンデータは産業の新しい「資源」であると考えられており、「ビッグデータ」と並び、新しいビジネスのシードとしても注目が集まりつつある (JETRO/IPA 2013 米国オープンデータの動向調査. <http://www.ipa.go.jp/about/research/20130830.html> 2014 年 3 月 4 日確認)。実際、現在の EU 圏内におけるオープンデータによる市場規模は 4.2 兆円と推察されている (林 2014)。

生態学と関連の深い生物多様性分野においても、データのオープン化への関心は国際的に急速に高まっており、2013 年には Semantics for Biodiversity (<https://semantic->

biodiversity.mpl.ird.fr/ 2014 年 3 月 4 日確認)、Biodiversity Informatics Horizons 2013 (<http://conference.lifewatch.unisalento.it/index.php/EBIC/BIH2013> 2014 年 3 月 4 日確認)、TDWG2013 (<http://www.tdwg.org/conference-2013/> 2014 年 3 月 4 日確認) などのカンファレンスでオープンデータのセッションが生まれ、同年 11 月には台湾でまさに生態・生物多様性情報のオープン化をテーマとした国際カンファレンス 2013 International Conference on Open Data in Biodiversity and Ecological Research (<http://taibif.tw/en/opendata2013/en/home> 2014 年 3 月 4 日確認)が開催された。生物多様性情報のオープン化は、生物多様性条約で定義された愛知ターゲット達成への貢献が期待されるなど (GBIF Secretariat 2013 Global Biodiversity Informatics Outlook, <http://www.biodiversityinformatics.org/> 2014 年 3 月 4 日確認)、様々な応用の可能性を秘めている。

日本では科学データのオープン化の動きは鈍いが、国際社会がオープンデータの潮流の中にあること、科学情報はオープン化の重点分野であること (「日本のオープンデータ憲章アクションプラン (案)」<http://www.kantei.go.jp/jp/singi/it2/cio/dai53/53gijisidai.html> 2014 年 3 月 4 日確認) から、今後、科学データのオープン化の動きが急速に進むことが考えられる。日本の生態学分野では、多量に死蔵されたデータが存在することが指摘されているが (新山ほか 2007; 石原ほか 2010)、これらを発掘・オープン化し、関連研究を進展させ、さらには研究だけでなく、保全や生態系管理等の実務における活用を促進させる好期が来ている。生態学者が「オープンデータ」の考え方やその発展の可能性を把握しておくことは、今後の生物多様性研究およびその応用に重要だといえよう。そこで本稿は、現在急速に広がりつつある「オープンデータ」についての概説を通し、国内の生物多様性分野におけるオープンデータ利用の積極化および、生態学データのオープンデータ化に向けた課題等を論じる。

オープンデータとは

オープンデータについて明確な定義があるわけでは無いが、オープンデータの普及に先進的に取り組んでいる Open Knowledge Foundation によれば、「一片のデータあるいはコンテンツが、誰もが自由に利用、再利用、配布できるもの」と定義され、1) 利用できる、そしてアクセスできる、2) 再利用と再配布ができる、3) 誰でも使えるという 3 点を全て満たすことが要件とされている (Open Knowledge Foundation 2012 Open Data Handbook

Documentation, Release 1.0.0. <http://opendatahandbook.org/> 2014 年 3 月 4 日確認)。重要な点として、出典の明示やライセンスの継承などの最低限の条件を満たせば、商用利用、データの加工、他の情報と組み合わせての再配布が全て可能なことが挙げられる。つまり、データ (ベース) の著作権 (コラム 2) に対する個別的な利用許諾が不要であることを意味する。

データを「オープンデータ」として公開する際には、わかりやすい形で、そのデータがオープンであることを宣言する必要がある。その際に使われるのが、オープン化に対応した利用条件、すなわち、自由に利用可能なことを明記できるライセンスである。オープンデータに付与されるライセンスを含めた知的財産権の問題については後述する。

オープンデータとフリーデータの違い

オープンデータという考え方が出てくる以前から、多くのデータが無償で公開されてきており、そのようなデータは、「フリーデータ」と呼ばれてきた。ここで、「オープンデータ」と「フリーデータ」は、必ずしもイコールでないことに注意していただきたい。なぜなら、一般にフリーデータと表現した場合、それはデータが無料で公開されていること、すなわち前述したオープンデータ条件のうち 1) あるいはその一部のみを満たしているデータを指すからである。言い換えれば、フリーデータは、2) 以下の条件である無償かつ自由な利用を保証するものではない。ここは重要な違いであり、利用者は注意する必要がある。多くのフリーデータは独自のライセンスを持ち、利用に際し、商用不可、教育目的のみ利用可、再利用して作成した情報を公開する際には事前に申請や許可が必要である等、厳密な条件がデータセットごとに決められているのが一般的である。たとえば、日本の自然史博物館標本情報のデータ集約サイト「サイエンスミュージアムネット」の利用条件は基本的に許可制であり (<http://science-net.kahaku.go.jp/> 2014 年 3 月 4 日確認)、Japan Long-Term Ecological Research (JaLTER) データベースは、利用を研究目的に限っている (<http://www.jalter.org/> 2014 年 3 月 4 日確認)。

オープン化の利点

データを「オープン化」することによる利点を考えてみる。データのオープン化は、それを活用した研究やサ

ービス、アプリケーションが生まれやすくなることと直結するため、それら成果の増加が大きな利点の一つである。生態学分野における例としては、GIS データの活用が挙げられる。かつて、生態学研究で GIS を利用するためには、研究に必要な基盤データ、例えば調査地のポリゴンデータ、調査ラインのベクタデータ等を全て自分で整備する必要があった。そのためには多くの労力、または金銭を要するため、GIS を利用する生態学者は限定的であった。しかし、ここ数年、国内で急激に GIS を利用する生態学者が増え、それを活用した生態学研究が増えている。そしてその多くは、環境省が公開している植生図 (<http://www.biodic.go.jp/trialSystem/top.html> 2014 年 3 月 4 日確認) や国土交通省が公開している国土数値情報 (<http://nlftp.mlit.go.jp/ksj/> 2014 年 3 月 4 日確認)、国土地理院が公開している基盤地図情報 (<http://www.gsi.go.jp/kiban/> 2014 年 3 月 4 日確認) を利用している (例えば Tsuji et al. 2011; Osawa et al. 2011, 2013)。これら地図情報は現在、インターネットを通じて無償で取得でき、研究を含めた様々な用途に活用することができる。こういった基盤情報が公開され、無料で入手できるようになったことと、GIS を利用した研究が広がったことは、極めて密接に関係していると言ってよいだろう。

ただし、上述の GIS データ類は全てが「オープンデータ」であるわけではなく、「フリーデータ」にカテゴライズすべきものもある。それらの間の大きな違いは、修正 (改変) や再配布の可否である。オープンデータの条件であり、もう一つの利点として挙げられるのは、この修正 (改変)、再配布可という点である。修正 (改変)、再配布可という条件は、具体的には、そのデータにミス等があった場合、迅速に発見し、修正 (改変) したものをすぐに再配布できるという利点を生み出している。一般にデータが巨大であればあるほど、その整理には労力がかかり、人為的なミスが発生する確率は高くなる。さらに、その再配布にも大きな手間がかかる。機械によって自動的に取得されたデータであっても、機械の故障やプログラムのバグ等によって不適切なレコードが混入することは起こりうる。データをオープンにすることで、結果的に多くの人間によるデータチェックが行われることになり、こういったデータのミスを迅速に発見することができる。加えて、修正した結果を再配布できることにより、データの修正の効率化が期待できる。これは同時に、そのデータを使ってさらに新しいデータを作ったり、他データと組み合わせた二次成果物を再配布することも意味する。これにより、データそのものの価値が向上し、データ作

成者の評価にもつながる。このような思想に基づいた類似例として、オープンソースソフトウェアの開発が挙げられる。オープンソースソフトウェアは、ソースコードを全て公開し、ユーザーらによってバグを発見、修正、機能の追加を実施することにより、開発の効率化や高機能の実現を目指している (例えばオープンソースの統計パッケージ R <http://cran.r-project.org/> や GIS ソフトウェア QGIS <http://www.qgis.org/ja/site/> とともに 2014 年 3 月 4 日確認)。基本的な考え方はデータもソフトウェアも同じであるといえよう。

さらなる利点として、データの収集・生成が一度のみで済むことも挙げられる。様々な分野で研究データが生成されているが、データの多くはオープン化されことなく、そのまま死蔵される。このため、同じデータをとるために、複数回の労力や資金が投入される場合がある。特に物理環境の観測データ等では、しばしばそういった状況が起こりうる。データがオープンにされることにより、同じデータの収集・生成は 1 回で済むようになり、効率化を図ることが可能になる。

オープン化の問題点

とはいえ、データのオープン化には問題も存在する。最も大きな問題は、データの権利、すなわち知的財産権に関するものであろう。データをオープンにすることと、著作権等、各種権利の非行使とは必ずしもイコールではないことに注意しなければならない。極論としては、オープン化のために全ての権利を放棄するという考え方もありうるが、権利放棄とは、データの公開者が利用者等から正当に得られる報酬等を全て放棄することを意味するため、公開者の立場として簡単には実施できない場合が多いだろう。よって、オープンデータにおける権利は、オープンデータとしての条件を担保できるライセンスによって管理することが、現実的かつ実効的と考えられている (権利の非行使を明確化するという考え方もある: コラム 3)。とはいえ、データごとに独自のライセンスを定める等、ライセンスの種類が多岐にわたると利用性を著しく損なう可能性があるため、共通したライセンスを利用することが望ましい。オープンデータに付与する共通ライセンスとして現在もっとも広く使われているのが Creative Commons ライセンス (以下 CC ライセンス) である。CC ライセンスは、民間非営利団体である Creative Commons (<http://creativecommons.org/> 2014 年 3 月 4 日確認) によって管理、更新されているライセンスのことで、日

本にも管理団体の支部としてクリエイティブ・コモンズ・ジャパンがある (<http://creativecommons.jp/> 2014 年 3 月 4 日確認)。CC ライセンスは、利用範囲がコントロール可能で、かつ利用者に対して余計な手続き等を求めないという点で大きな利点がある (コラム 4)。

オープンデータに限らず、公開されたデータを再利用する場合、その利用者がしばしば直面する問題として、精度の問題が挙げられる。生態学データを例に取ってみよう。多くの生態学者はフィールドワークによって自身で生態現象を観測し、データ化する。この場合は観測者と利用者が一致しているため、そのデータが持つ精度について利用者自身が責任を持つことができる。それにより、適用可能な解析手法等を自身によって判断できる。対してオープンデータを利用する場合、基本的に観測者と利用者は別となる。よって、観測者が想定しない利用が行われる場合も少なくない。この場合、その利用目的に必要な精度が確保されているかどうかについて、利用者が注意を払わなければならない。この問題は、例えば Species Distribution Model (SDM) (Guisan et al. 2013) を構築する際に起こりうる。地球規模生物多様性情報機構 (Global Biodiversity Information Facility : GBIF) のポータルサイトからは、生物の分布情報 (Occurrence records)、すなわち緯度経度等の位置情報が付与されたデータが取得できるが (GBIF portal : <http://www.gbif.org/> 2014 年 3 月 4 日確認)、そこに記載された位置情報は必ずしもピンポイント、すなわち付与されている緯度経度等の位置情報が、確実にそのデータを取得した地点を正確に示しているとは限らない。むしろ実際は、市町村、都道府県が特定できる精度の精度しかない場合が圧倒的に多い (例えば Osawa 2013 は GBIF から植物の観察データを公開しているが、町を 6 区画に分けた単位での位置情報のみ公開している)。こういった位置情報に数キロ程度の誤差を含むデータを利用して、10 m 精度の土地利用情報を説明変数とした SDM を構築しても、生物情報の誤差が土地利用情報の違いをキャンセルしてしまうため、その結果に意味はない。オープンデータを利用する際、利用者は自身の責任において精度の管理を実施しなければならない。同時にデータ公開者は、利用者のために精度情報等、データが持つ性質について詳細にメタデータ (コラム 5) に記載しなければならない。また、オープンデータであるが故の問題として、前述のようなデータに元々含まれる、または意図しない誤差とは別に、改変、再配布の際に意図的に誤ったデータが混入される危険性もある。データがオープンであることは、そのようなデータの検証

が可能であるということでもあるが、再利用に当たっては、そのような危険性に留意することも必要とされる。

オープンデータの現状

様々なデータのオープン化が進む中で、現在特に急速に進められているのが、政府の持つ様々な公共データをオープン化する「オープンガバメント」である。オープンガバメントは既に国際的な潮流となっており、アメリカの DATA.GOV (<http://www.data.gov/> 2014 年 3 月 4 日確認)をはじめ、20 ヶ国以上が国のオープンガバメントのためのポータルサイトを設置している。日本では、2012 年に「電子行政オープンデータ戦略」(内閣官房 2012 電子行政オープンデータ戦略 http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryu2.pdf 2014 年 3 月 4 日確認) を国家戦略として掲げ、オープンデータの流通の促進を産学官で進めていくことを目的としたオープンデータ流通推進コンソーシアム (<http://www.opendata.gr.jp/> 2014 年 3 月 4 日確認) が立ち上げられた。この中で、技術・ライセンス・活用の 3 分野について専門委員会が設置され、国内での技術的・法律的・文化的な課題について議論されている。実際、2013 年 12 月には、国内の公共データを集約し、オープンデータとして公開するためのデータカタログサイト DATA.GO.JP の試行版が公開された (<http://www.data.go.jp/> 2014 年 3 月 4 日確認)。DATA.GO.JP から取得できるデータには、すでに別の場所で公開されているが、改めてライセンスを CC-BY に変更されて登録されたデータも多い。この中には、環境省生物多様性センターの公開情報をはじめとした生物多様性情報も含まれる。2013 年 5 月に閣議決定された「世界最先端 IT 国家創造宣言」によれば、2013 年と 2014 年をオープンデータの集中取組期間と位置づけ、2015 年には他の先進国と同水準での公開を目標としている (<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryu1.pdf> 2014 年 3 月 4 日確認)。

もう一つ、めざましい動きを見せているのが、図書館や博物館などの社会教育分野である。その先駆けとなったのが、EU で 2008 年に公開されたヨーロッパアナ (Europeana, <http://www.europeana.eu/portal/> 2014 年 3 月 4 日確認) である。現在までに 600 万点以上の美術品・自然史標本・文献類の情報がパブリックドメイン (コラム 3) によって公開されている。著作権が切れた情報が中心だが、オープンデータとして公開されているものもある。アメリカでは、国内の主な博物館の全ての自然史標本の電子化と公開を目指した大規模プロジェクト Integrated

Digitized Biocollections (iDigBio, <https://www.idigbio.org/> 2014年3月4日確認)により、標本画像や情報のオープン化と、そのための様々な技術開発・情報学的研究が進められている。このような、図書館や博物館・美術館などの情報をオープン化する、あるいはオープン化における図書館の役割を考える動きを先導する団体が OpenGLAM (GLAM = Gallery, Library, Archives and Museum)である (<http://openglam.org/> 2014年3月4日確認)。2013年11月には、日本での OpenGLAM 活動の推進を目的とした団体 OpenGLAM JAPAN が立ち上げられた (<https://www.facebook.com/OpenGLAMjp> 2014年3月4日確認)。図書館情報学については、生態学データのオープン化に貢献する可能性も指摘されている (真板 2014)。

生物多様性の分野においてもデータのオープン化は広がりつつある。特に注目すべき例として、Pensoft (<http://www.pensoft.net/journals/> 2014年3月4日確認)の取り組みを挙げたい。出版社である Pensoft は、オープンアクセス誌 (コラム 6) として動物専門の Zookeys や植物専門の Phytokeys 等を出版している (<http://www.pensoft.net/journals/> 2014年3月4日確認)。これらの論文誌に掲載された論文そのものおよびデータは、全てオープンライセンスによってオープン化するのがポリシーとなっている (Penev ほか 2011 http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf 2014年3月4日確認)。さらに Pensoft は、分類学的な論文の出版とオープンな形でのデータ公開を促進するため、データ公開とデータベース連携を前提とした論文執筆システムを組み込んだ新たなデータペーパー (コラム 7) 専門雑誌「Biodiversity Data Journal」を 発 刊 し た (<http://biodiversitydatajournal.com/> 2014年3月4日確認)。Pensoft はデータの公開に際し、GBIF のデータプロトコルを採用していることもあり (<http://www.gbif.org/resources/2533/> 2014年3月4日確認)、生物多様性情報学の分野において最も注目されている取り組みの一つであるといえる。

生物多様性データの オープンデータ化に向けた国内の課題

日本では、生態情報学 (Ecological Informatics) を含む生物多様性情報学 (Biodiversity Informatics) が進んでいるとはいえない状況にあり (大澤・神保 2013)、生物多様性データのオープンデータ化も進んでいない。日本の生物多様性分野におけるオープンデータ化に向けた最大の課題は、データを公開する「文化」(GBIF Secretariat

2013 Global Biodiversity Informatics Outlook, <http://www.biodiversityinformatics.org/> 2014年3月4日確認) が根付いていないという点である。GBIF が発行している生物多様性情報概況 (Global Biodiversity Information Outlook : GBIO) では、生物多様性研究の推進にも、愛知ターゲットの達成にも、まず何より先に、生物多様性データがオープンデータとして共有され、オンライン上で自由に利用可能なリソースにするための基盤を構築し、そこにありとあらゆる関係者が協力することが当たり前になる「文化」を根付かせなければならないと言及している (GBIF Secretariat 2013 Global Biodiversity Informatics Outlook, <http://www.biodiversityinformatics.org/> 2014年3月4日確認)。日本における最大の問題は、個人、研究室、団体、プロジェクト単位での収集といったデータの規模に関わらず、原則としてデータは非公開が基本の「文化」となっていることである (武田 2011)。これはオープン化以前の問題である。実際、現在の日本において、自身の調査データ等を積極的に公開している生態学者は、ごくごく少数である。データを公開することが、当たり前のオプションとして捉えられるようにならないと、データのオープン化は進まないだろう。分子生物学分野では、すでにこのような「文化」が根付いており (コラム 1)、研究で生み出されたデータが集積され、集積したデータの再利用により研究や応用利用が発展し、それらがまた新たなデータを生み出す、という正のフィードバックが形成されている。データのオープン化に向けて、生態学分野においても、同様のフィードバックを形成すべきであろう。

そもそも、多くの生態学者は本来的にオープンにすべきデータを多数持っている。例えば科学研究費 (科研費)、環境省環境総合推進費、省庁からの委託プロジェクト等、明確に税金が予算元となっている研究プロジェクトにおいて生物多様性の研究が実施され、多くのデータが取得されている。細かく言えば、大学における研究活動も、少なからず税金が投入されている。これらは公金によって実施された研究である。オープンデータ憲章では、政府データは全て公開され、公的な利益として還元することを勧めており、税金によって得られた研究データも本来的には公開し、公的な利益に還元すべきであろう (オープンデータ憲章 http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html 2014年3月4日確認)。これについては、容易に受け入れがたい研究者も多いかもしれない。しかし、他の分野、たとえば気候変動や土地利用など、生態学分野においても基盤情報として使われる研究成果

が、当該分野の研究者により占有されていたらどうであろうか。関連研究は全く進まなくなるだろう。他の分野には公開を要求する一方で、自らの情報を公開しないというのは、利己的との批判を受けざるを得ない。また、報告書や論文として成果は公表しているという反論があるかもしれない。しかし、論文等にならなかったデータは一体どこに消えてしまったのであろうか？もちろん絶滅危惧生物の分布情報等、公開することで問題が起こりうるデータはあるが、その場合であっても、メタデータを公開することは可能であろう。メタデータの公開は、オープンデータ化への第一歩である。データの公開媒体やデータベース化の技術は既にある程度確立しつつあるので（大澤ほか 2011, 2012；大澤・神保 2013）、必要なのは意思のみである。行政データが積極的に公開されはじめていた昨今、論文にならなかった研究データも、プロジェクト等の終了等、積極的に公開していくことを考える時期が来ているのではないだろうか。巨大なプロジェクト等、ステークホルダーが多数にわたり、研究者の一人では難しい場合も多いであろうが、少なくとも議論は始めるべきであろう。データの公開、オープン化の利点が広く認識されるようになれば、地域の愛好家等の所有するデータについても、オープン化が進むことが期待される。これは、これまで研究の中では扱えなかった莫大な量のデータが使用可能になることで、新たな発見につながりうるものであり、生態学の進歩に是非とも必要なものである。読者の方々は、ぜひご一考いただきたい。

もちろんデータのオープン化それ自体にも、データ形式の問題やデータの置き場所の問題、それらの継続性といった技術的な問題はある。しかし、技術的な問題は多くの場合、時間とともに解決される。対して、文化の問題は待っていても解決されないのが普通である。まずは多くの生態学者が持つこの文化を変えること、これが生態学におけるオープンデータを推進するための最大の課題である。

おわりに

オープンデータの動向や現状について概説してきたが、オープンデータを取りまく状況は凄まじい早さで変化しており、数年と待たずに本稿の内容は陳腐化してしまうだろう。オープンデータという考え方は古くからあったが、それを実現可能としたのは情報科学の発展と、それらを利用した様々な成果物であり、そのさらなる発展とともに、オープンデータもさらに急速に発展していくと

予想される。情報科学関連分野は一般的に移り変わりが早いため、常に潮流を追いかけなければ、すぐに時代遅れになってしまう。しかし、現在の日本における生物多様性の分野では、情報科学を専門とした人間の職は皆無に等しい（大澤ほか 2013；大澤・神保 2013）。こういった状況では、そう遠くない未来、技術以前に、現在どんな考え方が世界の主流になっているのか、どんな研究が可能になっているのかさえ把握するのが困難になってしまうかもしれない。研究データの管理や公開は、学生やボスドクにちょっとただ働きさせれば片付くような単純な問題ではなくなったという指摘もある（真板 2014）。本稿で紹介した GBIF 等では、情報科学を専門とした人間が専門職として雇用され、あるいは生物多様性分野の研究プロジェクトに参画し、自身の専門知識を活かしてして最新技術や最新の考え方を取り入れている。対して日本では、情報科学専門家の雇用はおろか、プロジェクトに参画することすら滅多にない。日本の現状として、情報科学の専門家を擁する諸外国に生物多様性情報の分野で対抗していくのは難しい状況になりつつあると言える（大澤ほか 2013）。

オープンデータは、間違いなく生態学研究を推進してくれる新しい考え方である。特に意識していないかもしれないが、既に多くの方がオープンデータから恩恵を受けている。しかし、この考え方を有効に扱うためには、一定の知識、あるいは専門家の協力が必要である。日本の生態学コミュニティがオープンデータを有効に扱っていくために、オープンデータの「文化」を根付かせる必要がある。この実現に向けた議論を今後も進めていきたい。

本稿の著者らは、内容についてクリエイティブ・コモンズ・ライセンス 表示-継承 2.1 日本 (CC BY-SA 2.1 JP) で公開し、ライセンスに従う限り、商用を含む全ての利用・再配布を認めることをここに明示する。

謝 辞

オープンデータに関係する様々な方々と議論する機会を与えてくださった台湾 Academia Sinica、GBIF 台湾ノード TaiBIF の諸氏、特に Kwang-Tsao Shao 氏、Yu-Huang Wan 氏に謝意を表す。国立情報学研究所／LODAC の武田英明氏、亀田堯宙氏、Open Knowledge Foundation Japan 東修作氏、OSGeo 財団古川泰人氏には草稿をご校閲いただき、有用なコメントをいただいた。上述の方々のご協力に謝意を表す。本報告の一部は、ナショナル

バイオリソースプロジェクト (NBRP) の補助を受けた。

引用文献

- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12: Suppl 15:S2
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, 16:1424-1435
- 林 雅之 (2014) オープンデータ超入門. インプレス R & D, 東京
- 石原 正恵, 石田 健, 井田 秀行, 伊東 明, 榎木 勉, 大久保 達弘, 金子 隆之, 金子 信博, 倉本 恵生, 酒井 武, 齋藤 哲, 崎尾 均, 畠元 道德, 芝野 博文, 杉田 久志, 鈴木 三男, 高木 正博, 高嶋 敦史, 武生 雅明, 田代 直明, 田中 信行, 徳地 直子, 並川 寛司, 新山 馨, 西村 尚之, 野口 麻穂子, 野宮 治人, 日浦 勉, 藤原 章雄, 星野 大介, 本間 航介, 蒔田 明史, 正木 隆, 吉岡 崇仁, 吉田 俊也 (2010) モニタリングサイト1000森林・草原調査コアサイト・準コアサイトの毎木調査データの概要. *日本生態学会誌*, 60:111-123
- Ishihara M, Suzuki S, Nakamura M, Enoki T, Fujiwara A, Hiura T, Homma K, Hoshino D, Hoshizaki K, Ida H, Ishida K, Itoh A, Kaneko T, Kubota K, Kuraji K, Kuramoto S, Makita A, Masaki T, Namikawa K, Niiyama K, Noguchi M, Nomiya H, Ohkubo T, Saito S, Sakai T, Sakimoto M, Sakio H, Shibano H, Sugita H, Suzuki M, Takashima A, Tanaka N, Tashiro N, Tokuchi N, Yakushima Forest Environment Conservation Center, Yoshida T, Yoshida Y (2011) Forest stand structure, composition, and dynamics in 34 sites over Japan. *Ecological Research*, 26:1007-1008
- 真板 英一 (2014) 「研究データ公開」における人材と体制の問題—研究図書館の可能性—. *日本生態学会誌*, 64:81-86
- 三根 慎二 (2007) オープンアクセスジャーナルの現状. *大学図書館研究*, 80:54-64
- 新山 馨, 武生 雅明, 河原 崎里子 (2007) データベース化の功罪—森林動態データベース (FDDDB) を例に—. *日本森林学会誌*, 89:340-345
- Osawa T, Mitsuhashi H, Uematsu Y, Ushimaru A (2011) Bagging GLM: Improved generalized linear model for the analysis of zero-inflated data. *Ecological Informatics*, 6:270-275
- 大澤 剛士, 栗原 隆, 中谷 至伸, 吉松 慎一 (2011) 生物多様性情報の整備と活用方法—Web技術を用いた昆虫標本情報閲覧システムの開発を例に—. *保全生態学研究*, 16:231-241
- 大澤 剛士, 神山 和則, 桑形 恒男, 須藤 重人 (2012) Web APIを活用した個別データベースシステムの横断利用. *農業情報研究*, 21:1-10
- 大澤 剛士, 鎌内 宏光, 細矢 剛, 伊藤 元己 (2013) LTER, GBIFにおける国際的な生物多様性データベースの動向と日本国内の課題—国際ワークショップ参加報告—. *日本生態学会誌*, 63:269-273
- 大澤 剛士, 神保 宇嗣 (2013) ビッグデータ時代の環境科学—生物多様性分野におけるデータベース統合, 横断利用の現状と課題—. *統計数理*, 61:217-231
- Osawa T, Mitsuhashi H, Niwa H (2013) Many alien invasive plants disperse against the direction of stream flow in riparian areas. *Ecological Complexity*, 15:26-32
- Osawa T (2013) Monitoring records of plant species in the Hakone region of Fuji-Hakone-Izu National Park, Japan, 2001-2010. *Ecological Research*, 28:541
- Shao KT, Lin J, Wu CH, Yeh HM, Cheng TY (2012) A dataset from bottom trawl survey around Taiwan. *ZooKeys*, 198:103-109
- Takamura N, Nakagawa M (2012) The densities of bacteria, picophytoplankton, heterotrophic nanoflagellates and ciliates in Lake Kasumigaura (Japan) monitored monthly since 1996. *Ecological Research*, 27:839
- 武田 英明 (2011) 日本におけるLinked Dataの現状と普及に向けた課題. *情報処理*, 52:326-333
- Tsuji M, Ushimaru A, Osawa T, Mitsuhashi H (2011) Paddy-associated frog declines via urbanization: a test of the dispersal-dependent-decline hypothesis. *Landscape and Urban Planning*, 103:318-325

コラム 1

International Nucleotide Sequence Database (INSD)

INSD は、国際的に共有されている巨大な DNA 塩基配列データベースである。日本の国立遺伝学研究所の DNA Data Bank of Japan (DDBJ) センター、アメリカの National Center for Biotechnology Information (NCBI)、ヨーロッパの European Bioinformatics Institute (EBI) の 3 機関のデータベース群 (DDBJ, GenBank, ENL など) で構成されている。これらのデータベースに登録されている塩基配列情報は全て共有されており、どこで登

録された情報であっても、それぞれのデータベースから利用することができる。研究者が研究成果を発表する前には、これらの機関に塩基配列などを登録し、登録番号を得る必要がある。また、INSD の方針として、全ての登録情報をだれでも制限無しに自由に使うことができること、データを永久的に保存することなどが定められている (http://www.ddbj.nig.ac.jp/sub/insd_policies-j.html 2014 年 3 月 4 日確認)。分子生物分野では、解析で使用した情報の証拠 (パウチャー) と、その後の情報の自由な使用を担保する情報基盤と、公開と自由な利用を促進する文化が形成されているといえる。

コラム 2

著作権とオープンデータ

オープンデータは著作権と密接に関係している。著作権法によると、著作物とは「思想又は感情を創作的に表現したもの」であり、著作者は著作物を排他的に支配する財産的な権利、すなわち著作権を有することとなる (著作権法: <http://law.e-gov.go.jp/htmldata/S45/S45HO048.html> 2014 年 3 月 4 日確認)。そのため、他人の著作物を利用するためには、利用許諾を得る必要があり、その煩雑な手続きがデータの利活用を阻害して

いることが問題となっていた。他方、例えば「あいうえお順の電話番号データ」のように、創作性のない事実を羅列したデータそのものは著作物にあらず、誰かが著作権者として権利を主張することはできないのだが、どこまでが事実に関するデータで、どこまでが著作物であるかという線引きは困難である場合が多く、利用者を悩ませるケースが多々存在していた。そこで、利用許諾の手続きを簡略化、あるいは著作物であるなしの区別をすることなく利用可能とし、情報流通の促進を図るために生み出されたのが、オープンデータという考え方である。

コラム 3

パブリックドメイン

著作権は、著作物が創作された時点で自動的に発生する権利であり、極めて強い独占性を持つ権利である。その一方で、日本の著作権法には、その目的として「文化的所産の公正な利用に留意しつつ、著作者等の権利の保護を図り、もつて文化の発展に寄与する (原文ママ)」とあり、永続的に著作権が保護されることは利用の観点から望ましくないという面も持ち合わせている。そのため、一般に著作権はその有効期限が定められている (日本の場合は 50 年)。この保護期間が終了し、著作権が消滅したことをパブリックドメインという。

また、著作権が消滅する前に関係する権利の行使を放棄することを宣言する場合もあり、これもまたパブリックドメインと呼ぶ。本稿で対象とするのは主に後者が想定される。なお、特に著作権の放棄については様々な法的解釈が可能であり、利用が困難な場合もある。そのため、クリエイティブ・コモンズ (コラム 4 参照) では、CC0 というライセンスを作成することにより、パブリックドメインの利用を進めようとしている。CC0 については、クリエイティブ・コモンズ・ジャパンにより、日本語版ドラフトが公開されている (<http://creativecommons.jp/weblog/2013/11/5065/> 2014 年 3 月 4 日確認)。

コラム 4

クリエイティブ・コモンズ (Creative Commons)

クリエイティブ・コモンズ (以下 CC) は、様々な著作物がデジタル化されている現状を踏まえ、著作物の利用と流通を目指す活動であるとともに、その主体となる団体の名称である。著作権は著作者が持つ権利を保護する重要な権利だが、インターネット時代の著作物の利用 (共有や再利用) においては、コラム 2 で述

べたように、極めて不便である。そこで、利便性を高めるため、一部の権利のみを主張する目的で提案された。CC では「表示」、「継承」、「非営利」、「改変禁止」の四つの条件を設定し、それを組み合わせることによりクリエイティブ・コモンズライセンス (以下、CC ライセンス) を公開し (下図)、著作物の流通を図っている。なお、下図ライセンス一覧自体も CC-BY でライセンスされている。



図 CC ライセンス一覧



CC ライセンスのうち、オープンライセンスとされるのは、「表示 (CC BY)」と「表示-継承 (CC BY-SA)」の二つである。加えて、著作者が全ての権利の放棄を明示的に主張するための「利益放棄 (CC0): 法的な権利を望まない内容に関して放棄を明示」もオープンライセンスとみなされる (なお、パブリックドメイン (PD: コラム 3) は、著作物の知的財産権による保護が消滅した状態を指す)。CC ライセンスは最近数を増やしているオープンアクセスジャーナルで一般に採用されており (コラム 6) 研究論文のオープン化の際に付与するライセンスとしても知られている。また、CC ライセ

ンスを公共データに付与する動きも広がりつつあり、例えば福井県鯖江市では、全ての公開行政データに CC ライセンスを付与している (<http://www.city.sabae.fukui.jp/pageview.html?id=12765> 2014 年 3 月 4 日確認)。また、日本政府のオープンデータポータルサイトでも、「CC ライセンス 表示 2.1 日本」を採用している (<http://www.data.go.jp/> 2014 年 1 月 15 日確認) なお、CC ライセンスは、2013 年 12 月にバージョン 4.0 が作成され、国際対応とデータベースへのライセンス対応が改善された。新バージョンはまだ策定から時間がたっていないが、今後採用されるものと考えられる。

コラム 5

メタデータ (Metadata)

メタデータとは、「データのデータ」とも表現されるデータのことで、データセットそのものの説明を記述した情報のことである。例えば「いつ、どこで、誰が、どうやって取ったデータで、どんな形式 (フォーマット) で記述されたものなのか、作者は誰でどのような条件で利用可能か (ライセンス)」等が記述される。生態学のメタデータでは、データの観測日時、観測者、観測

場所、観測対象、観測目的等が記述されるのが一般的である。

データをオープンデータとして公表する際に、メタデータにデータの精度や収集方法、観測機器等を詳細に記述しておくことで、利用者はデータがもつ性質を正確に把握し、適切な利用を行うことができるようになる。逆に、メタデータに十分な情報がないデータは、利用性が極めて低くなり、オープンデータとしての利用価値を失ってしまいかねない。メタデータについての詳しい情報は、大澤・神保 (2013) を参照されたい。

コラム 6

オープンアクセスジャーナル

近年急増しているオープンアクセスジャーナルは、いわゆる一種の学術雑誌であり、基本的に電子ジャーナルかつ無償で誰でも読むことができるものである（三根 2007）。従来のジャーナルは読者が料金を支払う予約購読型モデルであるのに対し、オープンアクセスジャーナルにはいくつかの分類があり、例えば著者、読者とも無料である場合があるが、一般に著者が料金を支払う著者支払い・読者無料型モデルであることが多い（三根 2007）。オープンアクセスジャーナルの多くで CC ライセンスが採用されている（例えば PLOS ONE <http://www.plosone.org/static/license>; Hindawi

<http://www.hindawi.com/> 2014 年 3 月 4 日確認）。最近では、一般のジャーナルでも著者が所定の料金を支払うことで「オープンアクセス」として該当論文をオープン化することが可能である場合も多く、この場合も CC ライセンスが利用される（例えば Journal of Ecology <http://www.journalofecology.org/view/0/index.html> 2014 年 3 月 4 日確認）。オープンアクセスジャーナルも、データと同様の考えの下「オープン化」されている媒体の一つであると言えよう。日本では、国立情報学研究所の国際学術情報流通基盤整備事業（SPARC Japan、<http://www.nii.ac.jp/sparc/> 2014 年 3 月 4 日確認）によってジャーナルのオープンアクセス化を推進する事業が行われている。

コラム 7

データペーパー (Data Paper)

データペーパーは、データそのものの公表を目的とした新しい論文の形態であり、生物多様性分野で注目が集まっている（Chavan and Penev 2011）。生態学分野では、例えば森林モニタリングデータ（Ishihara et al. 2011）、湖沼モニタリングデータ（Takamura and Nakagawa 2012）等の長期観測データ（LTER）を公開する際に用いられる場合が多い。データペーパーの目的はあくまで「データ公開」であって、必ずしもオー

ペン化は想定していない点に注意が必要である。本文で触れた Pensoft の場合は、そのデータ公開ポリシーに基づき、データペーパーで公表したデータもオープン化がなされている（例えば Shao et al. 2012）。日本でも Ecological Research 誌がデータペーパーを発刊しているが（<http://www.springer.com/life+sciences/ecology/journal/11284> 2014 年 3 月 4 日確認）、2013 年 12 月現在で発表された論文数は合計 6 編と決して多くなく、これに掲載されたデータ自体も、オープンデータとして公開することは義務になっていない。