

特集

オープンサイエンスを支える 研究データ基盤

編集にあたって

村山泰啓 | 情報通信研究機構

林 和弘 | 文部科学省科学技術・学術政策研究所

オープンサイエンスの潮流

我が国におけるオープンサイエンスの潮流の変化は、2013年のG8科学技術大臣会合に端を発する。そこで「公的資金を得た研究データのオープン化」について共同宣言¹⁾がなされたのち、国内では内閣府が中心となって、本特集ゲストエディタも委員に加わり、国際動向と調整しつつオープンサイエンス推進の議論を行った。その成果が内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」の報告書「我が国におけるオープンサイエンス推進のあり方について～サイエンスの新たな飛躍の時代の幕開け～」²⁾として発表され、その後、我が国の統合イノベーション戦略³⁾において「オープンサイエンスのためのデータ基盤」が日本におけるイノベーションを生む重要な基盤になると言及され、これに基づいて国立研究機関のデータポリシー⁴⁾や国際的に信頼されるデータリポジトリのためのガイドライン⁵⁾などが公表されている。本特集のゲストエディタ2名はこれらの検討に直接関与してきたほか、国内外の会議での情報収集と我が国の関連活動情報の発信などにたずさわってきた。

また、第5期科学技術基本計画をうけて2021年4月以降の我が国の科学技術政策の基本方針となる科学技術・イノベーション計画（現状案）⁶⁾においても、オープンサイエンスへの対応を視野に入れつつ、デジタル化と

データの重要性が強調されており、今後も我が国の重要な政策と不可分に位置づけられているといえるだろう。

21世紀になってからのオープンサイエンスの本質は、インターネットの普及と電子情報基盤の社会利用の一般化により、学術情報の伝搬速度と配信コストが驚異的に変化し、広範囲な学術研究全般の方法論にまで影響する点にあると考えられる。特に、本特集でも着目している研究データについては、たとえば欧州委員会やOECD（経済協力開発機構）等のオープンサイエンス戦略においても論文と並ぶ重要な科学的生産物と位置づけられ、その取り扱いは科学政策上の大きな課題としてオープンサイエンスと不可分に議論されている^{7),8)}。

その原則的な考え方は、研究データをはじめとするさまざまな学術活動からのアウトプットを研究成果として位置づけ、その相互利用、可能な限りの積極的な公開を通じて、インターネット基盤や電子情報基盤の発展に伴う情報環境の変化に対応した、新たな時代の学術研究に対応しようというもの、といえる。もちろん、公開とともに自由な利用が可能になれば、研究成果の再現性・透明性の確保はもとより、既存研究から得られたデータによる新たな研究が可能となり（再利用）、さらには、これまでつながることの少なかった異分野間での横断的研究が加速し、これを通じて過去になかった科学技術イノベーションへの発展と経済効

【デジタルプラクティスコーナー】

各記事の概要のみ掲載しております。本文は電子版

<https://www.ipsj.or.jp/dp/contents/publication/46/S1202-index.html> をご覧ください。



果が期待できると考えられている。

社会における情報の記録・処理・伝達のテクノロジーという点においては、Gutenbergの活版印刷テクノロジーの産業的確立・普及を通じて世界最初の学術ジャーナルが1665年に出版されて以来⁹⁾、現代までのさまざまなイノベーションは印刷媒体の郵送や物理的な対面方式でのコミュニケーションをほぼ中心にして実現されてきたと言っても過言ではないだろう。

印刷メディアには数百年にわたって確立されてきた情報のハンドリング手法、社会技術（SNS等を指すのではない広義のソーシャルテクノロジー）が存在し、世界規模の機関間連携、メタデータ（書誌情報）の整備、法的整備、出版・配送をはじめとする周辺産業の成立など、そこには電子情報メディアとは異なる地平が広がっている（そうした状況の一部はたとえば過去の論文¹⁰⁾などにも見て取れる）。世界初の学術ジャーナル出版以来350年以上という、経済や政治とは比較にならないタイムスケールで情報を整備・管理・保存するスキル、人材、機関が存在してきた。

現代社会が享受する鉄道、内燃機関や再利用可能エネルギー技術、航空機、人工衛星、携帯電話などの社会に不可欠な科学技術は、ほぼすべて過去の紙媒体に記され適切に保存された知を活用することで生まれてきたイノベーションであることに思いを致すことは無駄ではないはずである。その知恵やスキルは50年・100年単位で社会の中で保存され、技術が衰退しないための人材育成から制度・法整備、製造・利活用エコサイクルの維持まで、有形無形の

工夫によって現代社会が成立していることを忘れてはならない。

現在議論が盛んなデジタルトランスフォーメーションが、学術や産業界で実現する場合、こうした社会になくはない技術、システム、制度やそれらの統合されたエコシステムを、インターネットをふくむ広い意味の電子情報基盤上で形成し、そこで新たな発見、発明、開発がなされて社会へ提供される時代をかたち作ろうというのであれば、それは非常に壮大なチャレンジであるとの見方もできよう（その一部はたとえば過去の報告書¹¹⁾などにも見て取れる。全体総括的な総説をご存じの方はご教示いただければありがたい）。もしそれが正しければ、紙媒体に代わって電子媒体が社会の基盤となっていくために乗り越えるべき膨大な課題がこれから私たちの前に将来にわたり広がっているであろうことは、頭の片隅に常におくことも重要であろう。

このような現状の中で、日本学術会議第24期では「オープンサイエンスの深化と推進に関する検討委員会」に多様な専門分野の方々が参画し、国際的な潮流に合わせて我が国におけるオープンサイエンスを進めるべきという提言がだされている¹²⁾。こうした動きを推進するために、情報科学や計算機工学、ソフトウェア工学などとともに、私たちはどういった新たな問題を解決すべきかを探るべき時に来ていると言ってもよいかもしれない。

【特集：オープンサイエンスを支える研究データ基盤】編集にあたって

本特集の論文について

本特集で採録された論文の紹介に先立ち、2編の関連解説論文について触れておきたい。

青木氏による解説論文「オープンサイエンスと研究データ管理の動向」は、デジタル技術に基づく新たな時代の研究データマネジメント（RDM）を、あえて対象学術分野の特定を避け、学術全体のためのプラットフォームと捉えて解説している。実際の研究適用時には分野ごとの特性、特徴の違いを考慮した実践が必要なのはもちろんであるが、これはデータ管理の変化やサイエンス（最も広い意味の学術、科学技術）の技術・文化・規範などの在り方の変化を暗に示唆しているともいえるだろう。

対照的に、高木氏による解説論文「統合データベースプロジェクトから学ぶこと」は、ゲノム研究・生命科学領域において著者ご本人が精力的にデータと格闘されてきた道のりをひもとき、データ管理、データベース構築、データ駆動型科学がいかに分野に依存し、人に依存し、組織に左右されるかといった、技術問題に落とし込む以前に横たわる課題とその解決の事例を示されている。サイエンス 2.0 と呼んでもオープンサイエンスと呼んでもよいが、デジタルテクノロジーの変革とともに学術の在り方の再考が求められ、今議論されている多くの課題の原型がすでに同氏のご経験の中で醸成されてきたことが伺われる。これらの教訓は我々にとっての今後の羅針盤として有益なものとなるだろう。

さて、以下は本特集において採録された各論文について概要を順次紹介したい。非常に多様な活動フィールド、組織、人々の視点から、共通する問題点もあれば、分野ごと・データごとに異なる事情が混ざり合い、運用上の問題と一言で片付けられない課題も多数ある。オープンサイエンス実現へ向けて、技術、文化、制度、学術規範、人的課題などを含めた総合的な研究課題が潜んでいることを意識して読

み取っていくことも今後有益であろう。

白井氏による論文「地球環境データベース—30年の歩みとこれから—」では、国立環境研究所（NIES）地球環境研究センターで30年以上にわたり継続的に整備されてきた「地球環境データベース」について、その時代とともに変遷する状況を、データベース公開、利活用促進、人的・文化的・組織的な観点から議論している。

能勢氏らによる論文「ジオスペース科学分野におけるデータ出版とデータ引用の現状およびそのプラクティス」では、地球周辺の宇宙空間を対象とするジオスペース科学、すなわち地上観測や人工衛星観測、計算機シミュレーションなどの手法を複合的に用いた電磁気的現象の研究において、2010年代半ばから推進してきた研究データへのデジタルオブジェクト識別子（DOI）付与のプラクティスを紹介している。多くの同分野研究者はまだ戸惑いがあるものの、学術出版社のデータポリシーの変化に伴い、今後データ引用の普及が加速していくと考えられ、データ出版数やデータ被引用数といった貢献度測定の可能性などを論じている。

川村氏らによる論文「データ駆動型農業に向けた研究データ基盤の構築」では、農業・食品産業技術総合研究機構（農研機構）における統合DBの整備、および、2020年5月よりこれと一体的に運用されているAI計算用スーパーコンピュータ「紫峰」にかかわるプラクティスを紹介している。ゲノムや品種、病害虫や環境に関する情報を対象として、FAIR原則に基づく共通メタデータを付けてカタログ化するだけでなく、複雑に絡み合ったデータ間の関係をRDF（Resource Description Framework）やProperty Graphなどを活用して表現する。これにより、ゲノム解析から育種、生産、加工・流通に至るサプライチェーン全体をカバーするデータ駆動型農業研究への道筋を論じている。

林氏らによる論文「JAIRO Cloud とコミュニティ主導のクラウドサービスの実現—」では、国立情報学研究所が世界でも類を見ない多数の



本編はこちら▶▶

利用機関（625 機関：2020 年 7 月末）に対して提供している JAIRO Cloud（機関リポジトリのクラウドサービス）について、その成功要因として、カスタマイズしやすい汎用リポジトリソフトウェア WEKO や、利用機関から構成されるコミュニティ主導の成長戦略を挙げるとともに、システムの安定運用や利用機関の獲得の過程の問題設定と解決等を紹介している。

大須賀氏らによる論文「情報学研究データリポジトリ IDR における研究用データセット共同利用の取り組み」では、データサイエンス研究の進展に不可欠な十分な規模のデータ利用を可能とするため国立情報学研究所の情報学研究データリポジトリ（IDR）で展開されているプラクティスを紹介している。産業界等と大学等の研究者の媒介、データセット共同利用、データ DOI 付与、利用実績や研究成果の状況を紹介するとともに、本活動を通じたデータセット共同利用の実現における課題発掘と理解の深化から対応への実践を議論している。

谷藤氏らによる論文「材料データプラットフォームシステム DICE における研究データフローの構築—実践と課題」では、材料分野でのデータ駆動型研究の進展を受けて物質・材料研究機構（NIMS）で 2020 年から開始された所内試験的サービス「DICE」における FAIR（Findable, Accessible, Interoperable, Reusable）なデータ流通基盤を紹介している。データを「つくる」「あつめる」「つかう」の基本コンセプトのもとで、マテリアル・インフォマティクスに不可欠な材料データベースや材料データリポジトリをオープンデータ基盤として再構築しており、オープンサイエンス時代に適合するための実践的取り組みと課題を考察している。

菊地氏らによる論文「CAS ベースの RDM 認証・認可機構の漸増開発とアセスメント評価」では、物質・材料研究機構における Research Data Management（RDM）に組み込まれた認証・認可機構（Central Authentication Service：CAS）の概略と設計上の変遷（認可管理との連携・名寄せ・多重化・API 管理）を概説

するとともに、Service Oriented Architecture（SOA）におけるセキュリティフレームワークの簡易アセスメントを実施するなどの評価・考察を試みている。

松波氏らによる論文「IoT データ収集システムのデータアーキテクチャ」では、物質・材料研究機構におけるデータ駆動型研究の進展に必要なデータ収集の仕組み、主に物質科学分野の計測・プロセスデータに関して、本来ネットワーク接続しない実験装置の IoT 化、計測データのメタデータ自動抽出から自動的データベース化といった設計指針（データアーキテクチャ）について論じている。

本特集での挑戦：研究成果の価値付けの新たな視点とデジタルプラクティスのデジタルトランスフォーメーション

本特集は、デジタルプラクティスというデジタル技術、情報科学などの実践に関する知恵や経験を共有し役立てるといったメディアにおいて、オープンサイエンスを進めるための萌芽的な試み、オープンサイエンスを指向しながら進めた従来の計算機基盤の修正・活用についての検討、またオープンサイエンス基盤そのものの開発などさまざまな試みが対象となっている。そこでは、計算機プラットフォームレイヤの試みから、プラットフォーム上でのデータコンテンツの管理実践の知、次世代へ残すべき共有情報資産としてのデータセットのあり方、これを広く活用するためのソフトウェアツールから人的・制度的枠組みまで、知的価値の高い情報の管理をどのように行うか、という「情報学」を幅広く捉えた試みがなされていると考えられる。

この編集過程においては、編集委員の間でオープンサイエンスと情報学の関係について、その位置づけや解釈において興味深い対話が生まれた。

まず、これまでのデジタルプラクティスにおいては、ポリシーを含む社会制度や、情報学としてのフレームはほぼ固定された中で、情報学の近傍のフレームでの

[特集：オープンサイエンスを支える研究データ基盤] 編集にあたって

実践が議論され、掲載されてきた。一方、オープンサイエンス時代においては、元となる社会制度自身もICTの変革やそれに伴う知識基盤の変化とともに大きく変わることが予察されており、予期せぬ形ではあるがCOVID-19がその予察を具体的なイメージに急速に落とし込み始めている。すなわち、今回のデジタルプラクティスを編集する上では、少なくとも研究者社会の制度や研究そのもののフレームを変え得るか、研究成果の価値付けを変え得るかどうかなどの「メタサイエンス」の観点が加わっている。

投稿された論文の中には、特に情報学の科学的視点から見た価値を見出せないと判断できそうなものがあったが、オープンサイエンスの実践に向けた何らかの価値を認められるものは積極的に採録した。つまり、従来の基礎科学的なもの、あるいは、その単なる応用研究とは明らかに違うベクトルによる価値付けにも我々は挑戦した。それは、新たな科学を生み出すフレームづくりのヒントとなる実学をどのように価値付けするかへの挑戦とも言える。

すなわち、本特集の編集作業は、単なる学術論文の査読を超えて、オープンサイエンス時代の新たな科学を生み出す活動の価値付けをどのように行うべきかの試行実験を行ったとも言えるだろう。これを一言でまとめれば、デジタルプラクティス自身のデジタルトランスフォーメーションも指向した結果として今回の特集は編集された。一見聞こえは良いが、実際は上に記した価値観が混在する中で一定の判断を下すために苦労することも多く、依然曖昧さが残るものである。

ジャーナル編集方針というのは、学問の在り方や評価の本質にかかわるがゆえに、編集にたずさわっ

たゲストエディタとしては、将来のデジタル基盤と広い意味での学術情報、さらには知的価値の高いデータ全般の管理や保全、世代を超えた学術知の管理と利用にとって有益な知をできるだけ多く読者に伝えるべく配慮をしたつもりである。

この取り組みをきっかけとして、情報学がさらに発展し、オープンサイエンス時代を支える知識基盤づくりが進展し、その知識基盤に基づく新しい科学研究が広く進展することを期待したい。

参考文献

- 1) G8 science ministers statement : London UK, 12 June 2013, <https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013> (Jan, 8 2020 参照)
- 2) 内閣府：国際的動向を踏まえたオープンサイエンスに関する検討会、我が国におけるオープンサイエンス推進のあり方について～サイエンスの新たな飛躍の時代の幕開け～(2015), <https://www8.cao.go.jp/cstp/sonota/openscience/index.html> (Jan, 8 2020 参照)
- 3) 内閣府：統合イノベーション戦略, <https://www8.cao.go.jp/cstp/tougosenryaku/> (Jan, 8 2020 参照)
- 4) 内閣府：国際的動向を踏まえたオープンサイエンスの推進に関する検討会、国立研究開発法人におけるデータポリシー策定のためのガイドライン(2018), <https://www8.cao.go.jp/cstp/stsonota/datapolicy/datapolicy.html> (Jan, 8 2020 参照)
- 5) 内閣府：国際的動向を踏まえたオープンサイエンスの推進に関する検討会：研究データリポジトリ整備・運用ガイドライン(2019), <https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/index.html> (Jan, 8 2020 参照)
- 6) たとえば、内閣府：科学技術・イノベーション基本計画の検討の方向性(案)(2019), <https://www8.cao.go.jp/cstp/tyousakai/kihon6/chukan/index.html> (Jan, 8 2020 参照)
- 7) たとえば、Commission High Level Expert Group on the European Open Science Cloud, : Realising the European Open Science Cloud (2016), https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf (Jan, 8 2020 参照)
- 8) たとえば、OECD : Open Science (2020), <https://www.oecd.org/science/inno/open-science.htm> (Jan, 8 2020 参照), および、OECD : Making Open Science a Reality, OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris (2015-10-15), <http://dx.doi.org/10.1787/5jrs2f963zs1-en>

論文誌 デジタルプラクティス「特集：オープンサイエンスを支える
研究データ基盤」はこちらでご覧いただけます（電子図書館）

https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_opensearch&index_id=10553





- 9) たとえば, Royal Society : Philosophical Transactions, 2020, <https://royalsocietypublishing.org/journal/rstl> (Jan, 8 2020 参照)
- 10) David, P. A. : The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution, Capitalism and Society : Vol. 3, Iss. 2, Article 5. DOI: 10.2202/1932-0213.1040 (2008)
- 11) OECD : OECD Science, Technology and Industry Scoreboard 2017: The digital transformation, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264268821-en> (2017)
- 12) 日本学術会議オープンサイエンスの深化と推進に関する検討委員会: 提言「オープンサイエンスの深化と推進に向けての提言」(2020), <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t291-1.pdf> (Jan, 8 2020 参照)

(2021年2月8日)

■村山泰啓 (正会員) murayama@nict.go.jp

1993年京都大学大学院工学研究科博士課程修了, 博士(工学)。同年, 郵政省通信総合研究所入所。アラスカ大学との北極域観測に関する日米国際共同研究などを経て, 2011年国際科学会議世界データシステム事業国際事務局ホストを担当, その後内閣府「国際動向を踏まえたオープンサイエンスに関する検討会」有識者委員, G7科学大臣会合オープンサイエンス部会共同議長, 日本学術会議国際サイエンスデータ分科会委員長, 日本地球惑星科学連合理事などを歴任。

■林 和弘 (正会員) kahayashi@nistep.go.jp

1997年東京大学大学院理学系研究科博士課程中退, 修士(化学)。日本化学会にて電子ジャーナル開発とオープンアクセスに携わり, 2012年から文部科学省科学技術・学術政策研究所(現)に着任。オープンサイエンス政策に資する調査研究と実践に取り組む。ユネスコオープンサイエンス諮問委員会委員および日本学術会議オープンサイエンスを推進するデータ基盤と, その利活用に関する検討委員会幹事等を歴任。

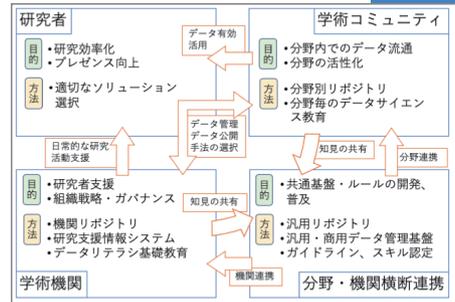
[特集: オープンサイエンスを支える研究データ基盤] 概要

1 オープンサイエンスと研究データ管理の動向



青木学聡 (名古屋大学)

研究成果を広く公開しその利活用を促すオープンサイエンスは, 新しい研究分野・イノベーションの創出をもたらすものとして, アカデミアの内外を問わず注目を集めている。また, 一連の研究活動に沿って実施される研究データ管理手法も, デジタル化の進展や研究活動の透明性向上等の点から, 見直しが進められている。本稿では, 2021年現在におけるオープンサイエンス, 研究データ管理の実現に向けた実践的な取り組みを多面的な視点で紹介する。



2 統合データベースプロジェクトから学ぶこと



高木利久 (富山国際大学)

生命科学は, ヒトゲノムの解読以来, ビッグデータを基盤としたデータ駆動型の研究に変貌を遂げつつある。このような背景のもと, 約15年前に我が国におけるこの分野のデータの共有・統合を目指す統合データベースプロジェクトが開始され, それを推進するためのデータベースセンターも整備された。本稿では, この15年にどういったことがあったのか, そこから得られた教訓は何か, 僭越ながらそれらを「10の教え」としてまとめたので紹介する。

4 バイオサイエンスデータベースセンターへの協力

バイオサイエンスデータベースセンター (<http://biocenter.nict.go.jp>) は, 様々な研究機関等によって作成されたライフサイエンス分野データベースの統合的な利活用を推進するために, 国立研究開発法人科学技術振興機構に設置されています。

同センターでは, 関連機関に積極的な参加を働きかけるとともに, 職員の交流, ポータルサイトの構築・運用, データベース統合の推進に関する研究開発, バイオデータベース統合の推進を目的として, ライフサイエンス分野データベースの統合化に向けて事業を推進しています。これによって, 我ら国におけるライフサイエンス分野の研究開発が, 広く研究者やコミュニティに共有・活用されることにより, 基礎研究や産業応用研究につながる研究開発を促すライフサイエンス分野の研究開発が促進されることを目指しています。

については, ライフサイエンス分野に関する論文発表等で公表された成果に関する生データの複製物, 又は構築した統合データベースの複製物について, 同センターの提供に御協力をお願いします。

なお, 提供された複製物については, 複製物の複製・改変等の必要な事項で御用命するものとします。また, 複製物の提供を受けた機関の求めに応じ, 複製物利用に当たって必要となる情報の提供にも御協力をお願いします。

55

をお願いすることがありますので, あらかじめ御了解ください。

また, バイオサイエンスデータベースセンターでは, 上記に関するデータについて, 個人情報保護法に配慮しつつ, ライフサイエンス分野の研究に係るデータの共有や利活用を推進するためにガイドラインを策定しています。