

オープンサイエンスと研究データ共有

大 向 一 輝

国立情報学研究所

Open science and research data sharing

Ikki OHMUKAI

National Institute of Informatics

In recent years, research data through advanced network environments and sensors has dramatically increased, and scientific research methods, called “fourth paradigm” and “data centric science,” have been attracting attention. From the viewpoint of research fraud prevention, the momentum of disclosing data on research results is increasing. These activities are widely called “open science.” To activate open science, it is necessary to promote daily data sharing inside and outside the research community. This paper introduces cases of open science on institutional and technical aspects, and discusses future prospects.

Key words: open science, research data, science community, data management plan, research platform, identifier

キーワード：オープンサイエンス，研究データ，科学コミュニティ，データ管理計画，研究プラットフォーム，識別子

1. はじめに

情報技術の進展やインターネットの普及は、人々のコミュニケーションやビジネスのあり方、さらには社会制度に至るまで多様かつ多大な影響を与えている。科学もまた例外ではなく、研究方法や成果発表のための手段が情報技術によって変化しつつある。近年ではこの変化の延長線上にある取り組みとしてオープンサイエンスが注目されており、その定義や実施方法、課題について活発な議論がなされている。本稿ではオープンサイエンスについて制度面と技術面の2つの観点から紹介するとともに、今後の展望について述べていきたい。

現時点においてオープンサイエンスの確たる定義は存在しておらず、百家争鳴の状況にある。欧州を中心としてオープンサイエンスに関する研修プログラムを提供している FOSTER プロジェクトでは、オープンサイエンスを取り巻く概念をオープンアクセス、オープンデータ、再現可能な科学、定義、評価、ガイドライン、政策、プロジェクト、ツールに分類し、それぞれに属する用語を整理した Open Science Taxonomy を公開して

いる（図1）。これらの概念や用語からもオープンサイエンスをめぐる状況の複雑さとステークホルダーの多様さを見て取ることができる。

本特集号の三浦（2018）による論考、また武田（2018）の整理にもあるように、オープンサイエンスは 1) 学術情報流通分野における論文入手コストの高騰化への解決策として進展してきたオープンアクセス、2) 政府や行政機関と市民との関係を再定義し、情報やデータの公開を進める公共オープンデータ、3) ネットワーク環境やセンサーの高度化によって可能になった、データを活用した研究手法を指す「第四の科学」あるいは「データ中心科学」、4) 専門家だけでなく市民が参加する科学活動としての市民科学、など出自の異なる取り組みを総称したものとして理解することができる。そのすべてを概観し、定義づけることは困難だが、本稿ではとくに 1) 2) の接点、すなわち公的活動としての研究成果のオープン化に関する活動を中心に取り上げる。

この観点におけるオープンサイエンスの定義は、政府・行政機関での議論の中から見つけることができる。例えば内閣府の総合科学技術・イノベーション会議に設置された「国際的動向を踏ま

Open Science Taxonomy

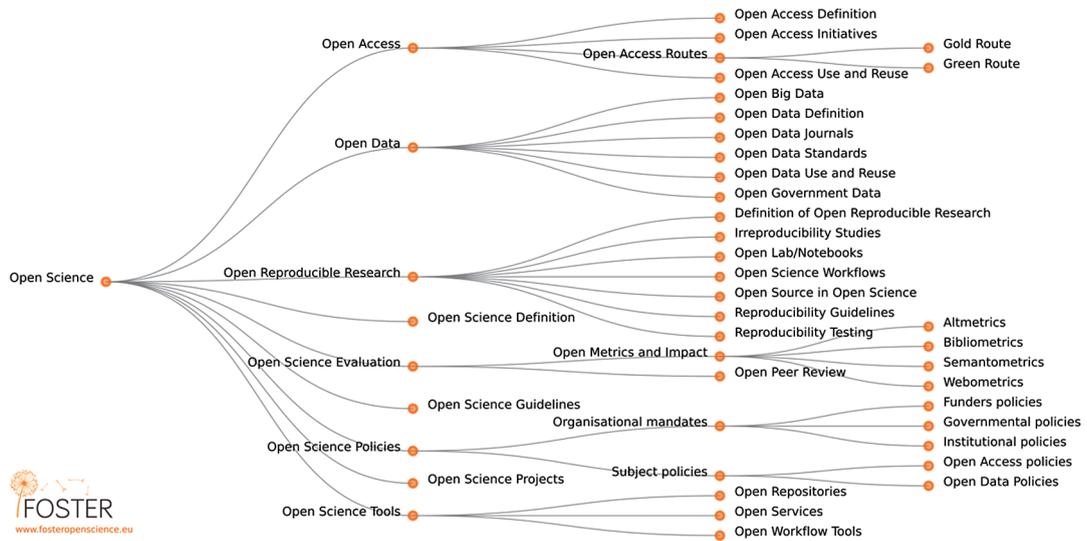


図1 オープンサイエンスタクソノミ (出典：<https://www.fosteropenscience.eu/foster-taxonomy/open-science>)

えたオープンサイエンスに関する検討会」の報告書 (内閣府, 2015) には下記の記述がある。

「オープンサイエンスとは、公的研究資金を用いた研究成果 (論文, 生成された研究データ等) について、科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開くとともに、効果的に科学技術研究を推進することで、イノベーションの創出につなげることを目指した新しいサイエンスの進め方を意味する。」

また2016年に策定された第5期科学技術基本計画 (内閣府, 2016a) には、「オープンサイエンスとは、オープンアクセスと研究データのオープン化 (オープンデータ) を含む概念である。(中略) また、オープンデータが進むことで、社会に対する研究プロセスの透明化や研究成果の幅広い活用が図られ、また、こうした協働に市民の参画や国際交流を促す効果も見込まれる。」と記載されている。

政府や行政機関によるオープンサイエンスの定義は、必然的に2)の公共オープンデータの影響を強く受けている。公共オープンデータは公的機関による調査や統計、報告書等を再利用可能な利用条件ならびに機械可読性の高い形式で市民に公開する取り組みである (大向, 2013)。2009年ご

ろから進められてきた公共オープンデータを巡る議論では、情報公開の目的として透明性 (Transparency)、参加 (Participation)、協働 (Collaboration) の3つのキーワードが挙げられており、前述の定義にもこれらの要素が含まれている。「透明性」は納税者に対する説明責任や知る権利への対応を指し、「参加」は情報公開による行政活動への興味関心や参加意欲の喚起を意味する。「協働」は民間セクターによるデータを活用した新規ビジネスの立ち上げなど、経済効果やイノベーションを狙ったものである。

研究活動を支える研究費の大半が公的資金であり、その原資が市民からの税金である現状においては、研究成果を公共オープンデータに準ずる形で公開すべきであるという主張は一定の説得力を持つ。このような状況の中で、政府や行政機関が科学コミュニティに対して半ば強制的にオープンサイエンスを課しているという見方があるかもしれない。他方で、科学コミュニティ自身にとっても透明性の確保は喫緊の課題であり、研究の過程で作成されたデータを第三者から確認可能な環境を構築することは不正防止に役立つ。また「参加」は学際的研究や市民科学の活性化、「協働」は公共オープンデータと同じく社会や民間への寄与と解釈すれば、科学コミュニティが能動的に

オープンサイエンスに取り組む理由になり得る。

今後のオープンサイエンスの成否は、公的機関によるトップダウンのアプローチと科学コミュニティによるボトムアップのアプローチが互いに呼応し、補完しながら進めていくことができるかどうかにかかっている。本稿では両者の観点から、各種の制度と技術に関する活動を紹介する。

2. オープンサイエンスの制度

2.1 トップダウンアプローチの動向

オープンサイエンスに関する制度は助成機関による整備が先行し、次いで政府レベルへと展開されてきた。最初期に研究データに対する取り組みを行ったのは米国の国立衛生研究所（NIH）である。NIHは2003年にデータ共有ポリシーを策定し、研究資金応募時のデータ共有計画の提出と、この計画に基づく成果の共有を義務化した。その後米国では2011年に国立科学財団（NSF）がデータ管理計画の義務化を行っている。データ共有計画ならびにデータ管理計画については後述する。

国際的には、2004年より経済協力開発機構（OECD）加盟国の科学技術担当大臣によって研究データの公開と利用に関する議論が開始され、2007年に「公的資金による研究データへのアクセスに関する原則及びガイドライン」が発表された（OECD, 2007）。続いて2013年にはG8科学技術大臣会合の共同声明において研究データのオープン化が合意され、さらに2016年にG7科学技術大臣会合で採択された「つくばコミュニケ」ではオープンサイエンスを分野横断的課題と位置づけ、国際的な検討体制の設置が明記された（内閣府, 2016b）。

これらの動きと前後して、各国内でのオープンサイエンス政策の整備が進み、英国研究会議（RCUK）による「データポリシーの基本原則」（2011年）、米国ホワイトハウスの科学技術政策局（OSTP）による「公的助成研究成果オープンアクセス指令」（2013年）など、公的資金による研究データを原則的にアクセス可能とするよう関係機関への働きかけが行われている。

欧州全体では、2014年から開始された研究支援プログラムである Horizon2020 においてオープンアクセスに関するガイドラインを策定するとと

もに、一部の研究プロジェクトを対象としてデータのアクセスと再利用を促進する Open Research Data Pilot (<https://www.openaire.eu/what-is-the-open-research-data-pilot>) を実施している。

民間助成機関の取り組みとしては、ビル&メリンダ・ゲイツ財団が2015年以降に助成を行った研究プロジェクトに対して、データを含む研究成果のオープンアクセス化を義務づけたことで注目を集めた (<https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>)。

日本では、これまでも文部科学省と日本学術振興会による科学研究費助成事業にて論文のオープンアクセス化が推奨されるなどの個別の取り組みが行われてきたが、近年の国際的な議論の展開に沿う形で先述の「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」の報告書や第5期科学技術基本計画におけるオープンサイエンス政策の明示化がなされた。さらには文部科学省の科学技術・学術審議会学術分科会学術情報委員会では「学術情報のオープン化の推進について（審議まとめ）」（文部科学省, 2016）を公表し、公的資金による研究成果のうち論文とエビデンスとしての研究データを原則公開とすべきこと、公開すべきデータの種類や形式、また非公開とすべきデータの判断は各分野の研究コミュニティによる検討が必要であることが記載された。これらを踏まえて、科学技術振興機構（JST）や日本医療研究開発機構（AMED）が助成する研究プロジェクトではデータ管理計画の提出義務化が開始されている。

2.2 ボトムアップアプローチの動向

前節で述べたトップダウンによるオープンサイエンス推進の一方で、科学コミュニティ側の対応も活発化している。国際科学会議（ICSU）では1966年より科学技術データ委員会（CODATA; <http://www.codata.org>）が設置されており、データの信頼性や検索性の向上に関する議論が行われてきたが、近年では研究データ管理と利用の促進についても議論の対象となっている。また、ICSUでは2008年に各国および各分野の研究データ関連組織を構成員とする World Data System（WDS; <https://www.icsu-wds.org>）を設立した。

2012年には研究者主導で研究データの共有に関する諸課題を議論する研究データ同盟（RDA;

<https://www.rd-alliance.org>) が組織された。RDA では研究者の他に図書館、政策担当者や助成機関を交えた多様な関係者が一堂に会して制度や技術についての30以上のワーキンググループを設けて検討を行っている。ワーキンググループにはメタデータや識別子の標準化など領域横断型の課題に取り組むものや、生命科学、人文学、農学といった分野特化型のグループが含まれる。またRDAと同様のステイクホルダーによって設立された非営利組織FORCE11 (<https://www.force11.org>) など、ボトムアップの取り組みは枚挙に暇がない。

出版社もデータ公開に協力する立場にあり、2016年から2017年にかけてSpringer Nature社やWiley社といった大手出版社がデータ共有ポリシーを発表し、論文投稿時のデータ公開が推奨されている。

2.3 データ管理計画

制度としてのオープンサイエンスを実施する上で重要な要素となるのがデータ管理計画 (DMP) である。データ管理計画のフォーマットは組織によって異なるが、典型的な例として英国Digital Curation Centre (2013) が作成したチェックリストには以下の項目が挙げられている。

1. 収集・作成対象データ
2. データ収集・作成方法
3. データに関する文書・メタデータ
4. 倫理的課題への対応
5. 著作権・知的財産権の管理
6. 研究期間中のデータ保存・バックアップ
7. アクセス方法・セキュリティ管理
8. 共有・保存等の対象データの選択
9. 長期保存計画
10. データ共有方法
11. データ共有に関する制約
12. データ管理責任者
13. 計画に必要なリソース

2.1 節で述べたように、研究助成の応募時にこれらの内容を含むデータ管理計画を策定し、提出することを求める機関が増加している。しかしながら、すべての研究者がこのような計画の策定に慣れているとはいえ、混乱が生じていることも事実である。そこで、大学図書館やリサーチアドミニストレータといった研究関連部局によって

データ管理計画の作成を支援する動きもある。先進的な大学や研究機関では、計画書の作成だけでなく、実際に得られたデータの管理や公開作業までを担当するデータライブラリアンあるいはデータキュレーターと呼ばれるスタッフを配置している例もある。

データ管理の望ましい形については、RDAなどの場で継続的に議論され、最終的にFORCE11 (2016) がそれらを取りまとめる形で提唱したFAIR原則がある。FAIR原則は、研究データが発見可能 (Findable)、アクセス可能 (Accessible)、相互運用可能 (Interoperable)、再利用可能 (Reusable) であることの頭文字を取っている。FAIR原則はオープンサイエンスのスローガンとして認知されており、個々の研究プロジェクトにおいてもFAIR原則に基づくデータ管理計画の策定が求められている。

3. オープンサイエンスの技術

3.1 研究のライフサイクル

これまで、制度面でのオープンサイエンスの取り組みについて紹介してきたが、これらはある種の理想状態を定義し、ルールや助成などの外的な力によって科学の方向をその状態に向かわせるアプローチであるといえる。一方で、現実の研究活動においては、データ管理計画の策定やその実施はコストの増加要因に他ならず、直接的なメリットやインセンティブがなければ広く普及することは難しい。そこで、情報技術の活用によって負担を軽減しつつ、研究者や支援者にとってのメリットを引き出すための様々な試みが行われている。

オープンサイエンスでは研究データの共有が謳われているが、その対象となるデータは最終成果物として論文中に取り上げられるものだけではなく、研究活動中のあらゆるプロセスにおいて中間生産物として多量かつ大量に生成されたものが含まれる。それらのデータのすべてについて共有や公開を行うためには、これまでの学術情報流通のように出版物の収集と管理、配布にとどまらず、研究のプロセス全体を支援するような情報環境を構築する必要がある。

英国情報システム合同委員会 (JISC) による研究プロセスと研究データの関係性を簡潔に示した

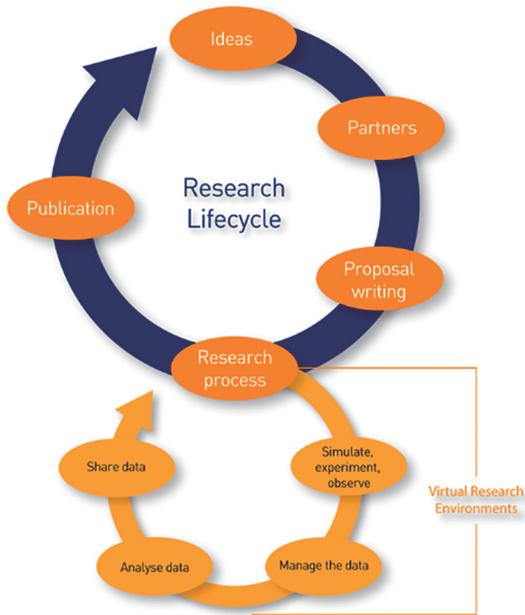


図2 研究のライフサイクル (出典：Research Lifecycle by David F. Flanders. CC BY-SA <https://www.flickr.com/photos/dff-jisc/4135521829/>)

図を示す(図2)。研究プロセスは着想から共同研究者の選定、提案書の作成を経て研究活動の実施に至り、成果物を発表することでサイクルが完結する。また研究活動の中ではシミュレーションや実験、観測によって得られたデータを管理し、分析、共有するサイクルが繰り返される。このような二重のサイクルを包括的に支援することが技術面でのチャレンジとなる。とくに、後者のサイクルについては、これまでは研究者が個別に行ってきたものであるが、研究環境の仮想化ないしプラットフォーム化によってサポートし、データの共有や公開へとつなげるという点で最も困難であり、また得られる利益も大きいものと予想される。

3.2 研究環境のプラットフォーム化

研究活動においてデータの共有を進めるためには、共同研究者が複数の組織にまたがる場合の円滑な共有方法や、データのバージョン管理、そして論文投稿時に提出するデータの長期保存方法などの課題がある。

これらの課題を解決するための研究環境のプラットフォーム化の取り組みで先行するのは欧州

である。欧州各国の研究データ管理機関が参加する EUDAT (<https://eudat.eu>) が2012年に設立され、研究者自身がデータを登録するための B2DROP、共有のための B2SHARE、保存のための B2SAFE、検索のための B2FIND といった一連のシステムが構築された。また組織横断型の分散コンピューティングを推進する European Grid Infrastructure (EGI; <https://www.egi.eu>) もクラウド環境やオンラインストレージを研究者向けに提供し、オープンサイエンスの支援を行っている。その後、2015年からは両者の活動がEUの支援の下で統合され、European Open Science Cloud (EOSC; <https://eoscipilot.eu>) と呼ばれるプロジェクトとして進行している。

米国においても大学および研究機関によって構成される National Data Service (NDS; <http://www.nationaldataservice.org>) が研究データ解析のためのクラウド環境やツールの共有を行っている。また University of Virginia の Brian Nosek が中心となって設立した非営利団体である Center of Open Science (COS; <https://cos.io>) は、研究プロジェクト単位でのデータ共有と公開を行うための Open Science Framework (OSF; <https://osf.io>) を提供するとともに、人材育成にも力を入れている。

オーストラリアでも検索サービス、クラウド環境、大規模ストレージの提供を担う各機関が連携し、Australian Research Data Cloud (<https://www.and-nectar-rds.org.au>) の構築を進めている。

国内の活動としては、筆者が所属する国立情報学研究所にてオープンサイエンス基盤研究センター (<https://rcos.nii.ac.jp>) を設立し、OSFをベースに研究者自身によるデータの保存、管理を行うデータ管理基盤、研究機関ごとに研究データを整理し公開するためのデータ公開基盤、それらのデータを集約し広く検索に供するためのデータ検索基盤の3種類のシステムを提供し、オープンサイエンスの活動を支援する予定である(図3)。

いずれの取り組みも、クラウド技術を用いた計算リソースやストレージの提供、組織の枠を越えたユーザ認証など多様な技術が必要であるとともに、ユーザである研究者にとって使いやすいインターフェイスを提供する必要がある。研究環境のプラットフォーム化はデータ管理の必要性から生じた側面が強いが、今後はプラットフォーム上で

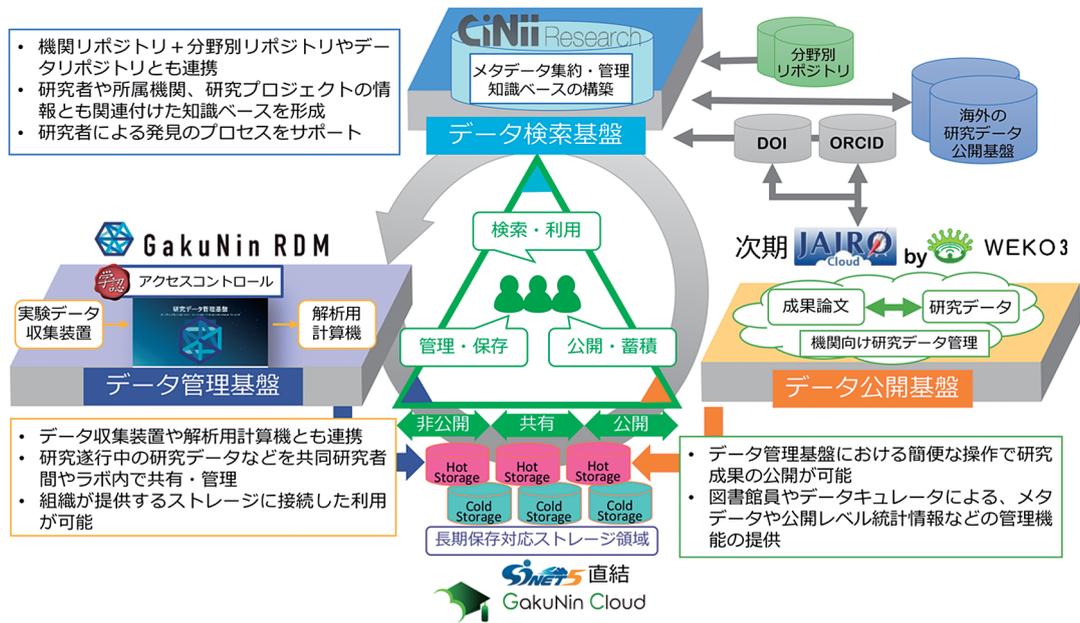


図3 国立情報学研究所によるオープンサイエンスの取り組み

解析プログラムを自動的に実行したり、その結果を保存することができるようになるなど、研究活動自体がプラットフォーム上で行われるようになる可能性があり、ビジネス領域としても注目を集めている。

3.3 オープンサイエンス時代の学術情報流通

従来の学術情報流通の枠組みの中では、書き手による出版や公開、ユーザへの提供、あるいは引用関係の把握や評価といった活動はすべて論文を媒介として行われてきた。本節では、今後研究データを含めた学術情報流通を実現するにあたって考慮すべき先駆的な事例を紹介する。

オープンサイエンスの源流のひとつであるオープンアクセスでは、出版社を介さずに文献を公開する方法として、研究者自らがウェブサイト等で発信するセルフアーカイブや、大学・研究機関などの組織単位で構成員の研究成果を公開する機関リポジトリが用いられてきた。とくに機関リポジトリは図書館などの組織に支えられていることから、研究データの収集や公開についても主導的な役割を担うことが期待されている。すでに機関リポジトリ上に研究データを格納し、公開している大学・研究機関も存在する。

欧州の機関リポジトリに掲載されたオープンアクセス文献の書誌情報を収集し、検索サービスとして提供する OpenAIRE (<https://www.openaire.eu/search/find>) では、2011年から研究データを対象とするプロジェクトを開始した。2013年にはCERNと共同で、誰もが研究データを自由に登録することができるリポジトリ Zenodo (<https://zenodo.org>) を公開している。

誰もが論文中の図表やデータを投稿し、共有することができる figshare (<https://figshare.com>) は個人が開発し、その後は Digital Science 社のサポートによって運営されている。figshare は大学や出版社と連携し、論文投稿時のデータの公開先としても用いられることで人気を博している。

データの公開においては、単にデータをアクセス可能な形でアップロードするだけでなく、そのデータを取得するために行われた実験の条件や、データそのものの特性について記載したデータデスクリプタを添付することでデータの信頼性が高まり、第三者による再利用の機会が増加する可能性がある。

データデスクリプタを論文と同様に扱い、査読プロセスを経て出版するデータジャーナルも発行されている。Springer Nature 社の Scientific Data

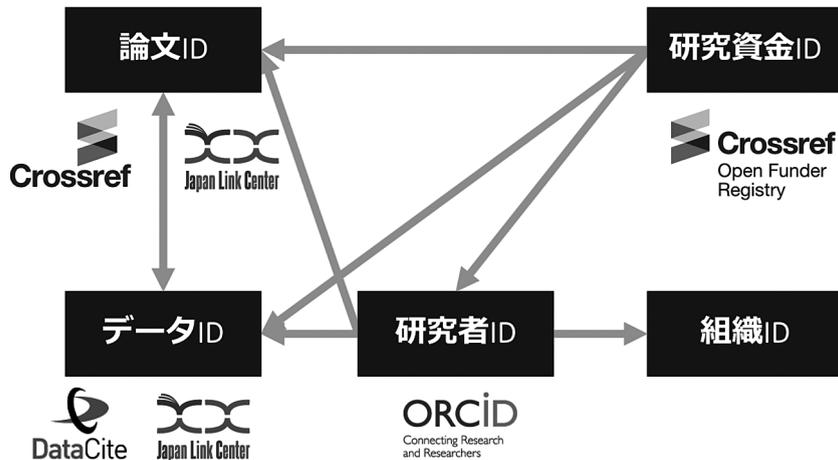


図4 多様な識別子のリンク（北本朝展（2016）研究データとオープンサイエンスに関する基礎的知識 https://japanlinkcenter.org/top/doc/160317_01kitamoto.pdf p. 19 に加筆）

(<https://www.nature.com/sdata/>) では、データデスクリプタを掲載するとともに、figshare との提携によってデータ本体の公開機能を提供している。また国立極地研究所が国内初のデータジャーナルとして Polar Data Journal (<https://pdr.repo.nii.ac.jp>) を発行した事例がある。

データあるいはデータデスクリプタは、論文から引用されることによって評価の対象となり、データを公開した研究者あるいは研究支援者に対するインセンティブとなる。

研究データを対象とした検索サービスは前述の OpenAIRE や、Elsevier 社の DataSearch (<https://datasearch.elsevier.com>) などが公開されている。研究データに付与されたメタデータやデータデスクリプタを対象とするもの、論文中の図表を対象とするものなどいくつかのタイプがあるが、今後研究者のニーズによって機能分化が進んでいくものと予想される。

3.4 研究データと識別子

研究データの管理や評価のために必須となる情報が識別子 (ID) である。従来から論文の識別子として Digital Object Identifier (DOI) が事実上の標準として用いられ、引用関係の把握にも役立ってきたが、近年研究データに対する DOI の付与を行うための環境が整備されている。

2009 年に設立された DataCite (<https://www.datacite.org>) は DOI の登録機関の 1 つとして研究

データへの DOI の付加サービスを提供している。また、科学技術振興機構、国立国会図書館、物質・材料研究機構、国立情報学研究所が共同で運営するジャパンリンクセンター (JaLC; <https://japanlinkcenter.org>) が国内資料の DOI 登録機関として論文だけでなく研究データへの DOI 付与を認めている。このような環境の中で、論文が研究データ DOI を参照・引用する文化が醸成されることが期待される。

情報自体に与えられる DOI の他に、研究者や研究支援者を対象とした識別子が ORCID (<https://orcid.org>) である。ORCID は人名の曖昧性を排除したうえで個人を特定するための 16 桁の数字であり、研究者自身が取得する。学術出版社の一部では論文の投稿時に筆頭著者ならびに共著者全員の ORCID の登録を義務化している例もある。今後、DataCite や JaLC にて研究データ DOI の取得時に ORCID の入力が行われるようになれば、研究者とその成果物としての関係性が明示化され、結果として評価につながるインセンティブとなり得る。

他にも、研究者が所属する組織や助成機関の ID、研究プロジェクトの ID などが続々と提案されており、それらが相互に紐付けられることで個々人の活動の成果や組織としての投資効果などを把握することが容易になる (図 4)。とくに研究組織や研究支援者にとっては評価基準となる指標が明確になることは多大なメリットがあり、今後

の科学の進展のためにも識別子の整備は重要な要素となるだろう。

4. 研究データ共有の萌芽的事例

ここまで研究データ共有を取り巻くトップダウンならびにボトムアップの活動を紹介してきた。現時点において、そのすべての提言内容やルール、技術を盛り込んだ事例は筆者の知る限り存在しないが、それぞれの研究分野において規模の大小を問わず着実に実施されている取り組みを見つげることができる。本節では国内の事例を紹介する。

天文学や地球科学をはじめとして、人工衛星や大型望遠鏡など大規模な装置を用いた観測が必要な分野では、装置ならびにその装置から得られるデータを共有することは研究活動を行ううえでの前提条件である。1989年に打ち上げられた磁気圏観測衛星「あけぼの」は、当初オーロラに関する物理現象の解明を目的として観測が行われたが、26年間に渡る運用の過程で得られたデータが、長期的な観測を必要とする太陽活動の解明にも役立つことが明らかになった。「あけぼの」の観測データが研究プロジェクト内外に広く公開された結果、このデータを利用した査読論文は300件を超え、現在に至るまで順調に増加している(常田, 2015)。また近年ではプロジェクト外の外国人が第一著者である論文の割合が30%を超えるなど、国際的なデータ共有が実現されている。

生命科学も研究データ共有が進んでいる分野の1つである。国内ではバイオサイエンスデータベースセンター(NBDC)が、公的資金を用いた研究プロジェクトで作られたヒトに関するデータを集約したNBDCヒトデータベース(<https://humandbs.biosciencedbc.jp>)を提供している。このデータベースは国立遺伝学研究所のDNA塩基配列データベースや大阪大学蛋白質研究所のタンパク質データベースと連携し、利便性を高めている。一方で、個人情報が含まれるデータが大量に存在することから、一律的に公開されるのではなく、情報の内容によって制限を行うなど柔軟な対応がなされている。

情報学分野では、大規模なデータが商用のインターネットサービス側に存在することが多く、研

究利用のためには個別に共同研究契約を締結する必要があった。国立情報学研究所では情報学研究データリポジトリ(<http://www.nii.ac.jp/dsc/idr/>)を設置し、商用サービス各社からのデータ受け入れと研究者への仲介機能を提供している。扱われているデータの中には、知識共有サービスの質問文・回答文のテキストデータや、動画共有サイトのコメントデータなどコミュニケーションに関わる情報が含まれており、心理学分野においても有用であると思われる。また、これらのデータを利用した研究の発表会を定期的実施することで、コミュニティの形成とデータ提供者に対する理解促進を図っていることが特徴的である。

人文学では、国文学研究資料館が30万点の古典籍を対象としたデジタル画像の作成と、新日本古典籍総合データベース(<https://kotenseki.nijl.ac.jp>)におけるオープンデータとしての公開を進めている。これらはすべて2次利用可能な条件で配布され、貴重資料を扱う他機関のデータベースの運用方針にも影響を与えている。一方で、情報・システム研究機構の人文学オープンデータ共同利用センターなどを中心として画像を活用した研究が進められており、機械学習を用いたくずし字の解読や、クラウドソーシングによるテキスト化など従来の枠組みに囚われない研究手法が提案されている。

5. オープンサイエンスの今後と人的支援

本稿では制度と技術という2つの観点からオープンサイエンスの動向について述べてきた。オープンサイエンスは2010年代におけるまったく新しい潮流であると主張することも可能であるが、一方で「巨人の肩の上に立つ」という言葉に示される科学の方法論を現代の情報環境の中で実践する試みにすぎないと捉えることもできる。

オープンサイエンスの概念とは無関係であっても、これまでも一部の研究者からは研究の過程で蓄積されてきた資料、例えば写真や動画のデータベース化と公開を行いたいという要望は存在してきた。そして、主にコスト面の問題から実現に至らず、結果として貴重な資料やデータが散逸したり、研究者の異動や引退に伴って見つからなくなるなどの状況が多数生じているものと思われる。

科学の多様化と大規模化に伴って、一研究者の努力でデータの共有を行うことが困難になっている側面は否定できないが、研究者単独ではなく組織や研究支援者との協力関係を強めることでこの課題に取り組むべきであろう。

例えば、京都大学では学内の研究データ管理基盤の整備が進められており、アカデミックデータの適切な蓄積・共有・公開および長期保存を行うためのデータマネジメント環境の調査研究を行うとともに、多様な研究分野の研究者とリサーチアドミニストレータ、図書館、博物館、文書館などの支援組織の関係者が共同でデータ管理を目指している (<http://www.cpier.kyoto-u.ac.jp/about/acd/>)。

その他にも、大学図書館が中心となって設立されたオープンアクセスリポジトリ推進委員会では、オープンサイエンス支援のコミュニティ形成と人材育成を目的として研究データ管理トレーニングツールの開発とオンライン教材としての提供を行っている (<http://www.nii.ac.jp/service/jmooc/rdm/>)。

今後はこういった人的支援の下で、メタデータやIDの付与を含めた情報管理、データのオープンアクセスを可能とするための権利処理とライセンスの付与、データベースなどの情報基盤の共有を行うためのマッチングや資金獲得を通じて持続可能な研究データ管理と公開を行うことが期待される。

文 献

- Digital Curation Centre (2013) Checklist for a Data Management Plan. v.4.0 <http://www.dcc.ac.uk/resources/data-management-plans> (2018年2月16日)
- FORCE11 (2016) FAIR Data Principles <https://www.force11.org/group/fairgroup/fairprinciples> (2018年2月16日)
- 文部科学省 (2016) 学術情報のオープン化の推進について (審議まとめ) http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm (2018年2月16日)
- 三浦麻子 (2018) 心理学におけるオープンサイエンス心理学評論, 61, 3–12. https://doi.org/10.24602/sjpr61.1_3
- 内閣府 (2015) 「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」報告書 http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_1.pdf (2018年2月16日)
- 内閣府 (2016a) 第5期科学技術基本計画 <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf> (2018年2月16日)
- 内閣府 (2016b) G7茨城・つくば科学技術大臣会合 つくばコミュニケ http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html (2018年2月16日)
- OECD (2007) OECD Principles and Guidelines for Access to Research Data from Public Funding <https://doi.org/10.1787/9789264034020-en-fr> (2018年2月16日)
- 大向一輝 (2013) 日本におけるオープンデータの進展と展望 情報管理, 56, 440–447. <https://doi.org/10.1241/johokanri.56.440>
- 武田英明 (2018) ORCIDとオープンサイエンス 平成30年 第17回分類学会連合公開シンポジウム (2018年1月6日・国立科学博物館) 講演資料 <https://www.slideshare.net/takeda/orcid-85786283> (2018年2月16日)
- 常田佐久 (2015) 磁気圏観測衛星「あけぼの」の運用終了について http://www.jaxa.jp/press/2015/04/20150410_akebono_j.pdf (2018年2月16日)