

わが国におけるデータシェアリングの
あり方に関する提言
(別添資料1)

ステークホルダーにおけるデータシェアリングの動向

1. 諸外国のステークホルダーにおけるデータシェアリングの重要性認識

近年、諸外国の政府組織、主要な研究資金配分機関といった重要なステークホルダーにおいて、研究データのオープン化、共有に関する各種の声明やポリシーが公開されている。各組織によって研究データのオープン化、共有目的は様々であるが、いずれの組織も重要な目的意識を持ってデータオープン化、共有を推進している。

以下に主要組織の例を紹介する。尚、その他の機関については、別添参考資料「研究データ共有ポリシーに関する調査」にて報告する。

(1) OECD 科学技術政策委員会閣僚級会合宣言¹

2006年12月、OECD加盟国において“OECD Principles and Guidelines for Access to Research Data from Public Funding(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)”が承認された。研究データへのオープンアクセスを実現し、その広範な利用を促すことが、研究の質と生産性の向上につながるという認識のもとに、「公開性」、「透明性」、「遵法性」、「公的責任」、「専門性」、「知的財産権の保護」、「相互運用性」、「品質とセキュリティ」、「効率性」、「説明責任」という10原則に則って、公的資金を活用した研究データの利用体制を整備していくべきであると謳っている。

(2) G8 科学大臣及びアカデミー会長会合共同声明

2013年6月、G8科学大臣及びアカデミー会長会合にて発せられた共同声明において、科学研究のオープンデータが言及されているが、その声明では科学的発見、イノベーションの創出、研究成果の幅広い可用性確保、科学の透明性確保、科学への国民参画を加速させるために科学研究データのオープン化を確約している。

(3) アメリカ合衆国大統領行政府 科学技術政策局(OSTP) 公的助成研究成果 OA 指令³

2013年2月、OSTPは年間1億ドル以上の研究開発費を有する研究助成機関に対し、2013年8月末までに“Increasing Access to the Results of Federally Funded Scientific Research(論文と科学データへのアクセス拡大計画)”を策定することを指示した。公的助成研究のデジタルデータセットへのアクセス拡大政策により、資源や労力を研究に集中させることが出来、データの再利用による研究投資の影響力や説明責任を最大化させ、キュレーション、保存、分析、可視化に関連する革新的サービスの経済市場形成と雇用創出を向上させるとしている。

¹ OECD Principles and Guidelines for Access to Research Data from Public Funding
<http://www.oecd.org/sti/sci-tech/38500813.pdf>

² G8 Science Ministers Statement <https://www.gov.uk/government/news/g8-science-ministers-statement>

³ Increasing Access to the Results of Federally Funded Scientific Research
https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

(4) 米国国立衛生研究所(NIH) データ共有ポリシーと実施要綱⁴

2003年10月、NIHはData Sharing Policyを制定し、研究資金公募時におけるデータ管理計画(Data Sharing Plan)の提出を義務化した。本要綱では、NIHの助成研究によるデータの共有理由として、「科学研究のオープン化の強化」、「分析や見解の多様化の奨励」、「新たな研究の推進」、「新規又は代替仮説と分析方法の検証」、「データ収集方法や手段に関する研究の支援」、「研究者育成」、「最初の研究者達が想定しえなかった課題の模索」、「複数のソースからのデータを組み合わせ、新しいデータセットの作成」を挙げている。

(5) 欧州委員会(EC) 科学技術・イノベーション政策 HORIZON2020 研究データ管理ガイドライン⁵

2014年1月より、FP7の後継フレームワークプログラムであるHORIZON2020が開始された。本ガイドラインでは、より深くより幅広い科学出版物及び研究データへのアクセスが、下記目的の助けとなるとしている。

- ① build on previous research results (improved quality of results)
これまでの研究結果の上に立脚する(結果の品質向上)
- ② foster collaboration and avoid duplication of effort (greater efficiency)
協力を奨励し、作業の重複を避ける(効率化)
- ③ accelerate innovation (faster to market = faster growth)
イノベーションを加速する(より早く市場へ=より早い成長)
- ④ involve citizens and society (improved transparency of the scientific process)
市民と社会が関与する(科学的プロセスの透明性の向上)

尚、欧州委員会(EC)では、Science2.0の概念について、科学研究手法のパラダイムシフトとして科学のプロセスを変革し、科学研究成果の増加とアクセス、科学的発見への信頼性、新たな研究手法としてのデータ集約型科学、シチズンサイエンス促進の観点から、政策としてどう介入するか検討を始めている。

(6) 英国研究会議(RCUK) Common Principles on Data Policy⁶

2011年、RCUKは「公的助成を受けた研究データは公益財であり、できる限り制限なく、適時にまた知的財産を害することのないよう責任ある方法でオープンに利用できるようにする」という基本原則のもと、“Common Principles on Data Policy”を公表した。本ポリシーは、政府助成研究成果に対する透明性の確保と統一的研究基盤の整備を目的として、研究データのオープン化に関する英国内各研究会議に共通の原則を定める包括的枠組みと位置づけられている。この枠組みに沿って、下記参考のように、英国における多くの研究助成機関ではデータポリシーが発行されており、データマネジメントおよびシェアリングの計画書提出を要求しているものが

⁴ NIH Data Sharing Policy
http://grants.nih.gov/grants/policy/data_sharing/

⁵ Guidelines on Data Management in Horizon 2020
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

⁶ Common Principles on Data Policy
<http://www.rcuk.ac.uk/research/datapolicy/>

多い。

■参考:英国の研究資金助成機関におけるデータポリシーの概要

● Full Coverage ○ Partial Coverage ○ No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	○	○	●	○	○	○
BBSRC	●	●	●	●	●	●	●	●	●	○	●
CRUK	●	●	●	●	●	●	●	○	●	○	○
EPSRC	●	●	●	○	●	●	●	○	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	○
MRC	●	●	●	●	●	●	○	○	●	○	○
NERC	●	●	●	●	●	●	●	●	●	●	○
STFC	●	●	●	●	●	●	●	○	●	○	○
Wellcome Trust	●	●	●	●	●	●	●	●	●	○	●

出典: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

(7) Research Data Alliance⁷ (研究データ連盟: RDA)

2012年8月に、アメリカ国立科学財団(NSF)、欧州委員会、オーストラリア国立データサービス(ANDS)の出資により創設された国際的組織である。政府機関の支援を受けているが、その議論は研究者主導で行われる。研究データの共有と交換を通じて「研究者とイノベーターが技術、分野、国を超えてデータをオープンに共有し、社会の大きな課題を解決するため、データのオープンな共有を可能にする社会的、技術的架け橋を築く」ことを目標としており、この実現のためにワーキンググループ等においてインフラ、ポリシー、プラクティス、規約等のグローバルルール開発と採択を行い、データシェアリングの「How-to」を議論する場として、運営されている。年2回のRDA総会開催とワーキンググループを中心に12ヶ月~18ヶ月で様々なテーマの議論を行い、テクニカルペーパーを策定するなど、議論のスピードが早い。また、米国NSFを始めとする政府系機関、国際組織や大学などの関係者がRDAに参加しており、RDAにおける検討結果が世界への波及効果が高いことが特徴としてあげられる。

⁷ Research Data Alliance <https://rd-alliance.org/>

(8) *World Data System*⁸ (WDS)⁸

2008年10月、国際科学会議(International Council for Science:ICSU⁹)において設立された、科学研究で利用されるデータを扱う個々のデータセンターやデータ関連組織を構成員とする国際組織である。品質管理された科学データの長期的な保全と提供によって、ICSUが推進する科学研究事業を支援することを主目的とする。自然科学系各分野はもとより、広く人文社会科学系までを含んだ幅広い領域におけるデータを対象とし、最新の情報科学研究の成果を取り入れたデータの保全・利用システムの構築や、発展途上国における情報格差の解消を目指す。日本では、日本学術会議の情報学委員会サイエンスデータ分科会の下部組織として、「WDS小委員会」が組織されているほか、独立行政法人情報通信研究機構内に国際プログラムオフィス(International Programme Office)を設置している。

(9) *Committee on Data for Science and Technology (CODATA)*¹⁰

国際科学会議(International Council for Science:ICSU)に設置された分野横断型の国際委員会で1966年に創設された委員会であり、米国、英国、フランス、ドイツ、カナダ、日本など23ヶ国が加入しており、科学技術データの管理と利活用を促進し、国際的な科学を強化し、社会の利益に還元することを目的としている。日本では、日本学術会議の情報学委員会サイエンスデータ分科会の下部組織として、「CODATA小委員会」が組織されている。

⁸ World Data System <https://www.icsu-wds.org/>

⁹ International Council for Science <http://www.icsu.org/about-icsu/about-us>

¹⁰ Committee on Data for Science and Technology <http://www.codata.org/>

■参考:海外の主要研究資金配分機関におけるデータ共有対象データ

主要機関におけるデータ共有の対象データは以下のように規定されている。但し、いずれの組織に おいてもプライバシー保護、機密情報、知的財産権、国家安全保障等に係る公開制限について別途規定されている。

機関名	対象データ
OSTP (米国大統領 行政府 科学 技術政策局)	政府機関が全部あるいは部分的に助成する機密扱いに区分されない研究から得られるデジタル形式の科学データ。 デジタルデータの定義: ・学術論文の裏付けとなるデータセットなど研究結果を立証するのに必要な科学界で共通に受け入れられるデジタル的に記録された事実に基づくデータ ・実験ノート、準備分析、論文の原案、将来の計画、査読報告書、同僚との通信、物理的対象物は含まない
NIH (米国国立衛 生研究所)	・研究を目的とする最終研究データ(科学コミュニティが研究結果を文書化しサポートするのに共通に必要とする記録された事実。サマリー表などではなくその元となるデータ) ・NIHが助成する基礎研究、臨床研究、調査等。特に複製が不可能なユニークデータ(莫大な費用がかかるため複製できない大規模調査、自然災害、稀な人口群に関する調査など)については特に重要 ・助成期間中、年 \$ 500,000 以上の直接経費を要求する申請 ・2003 年 10 月以降の申請
NSF (米国立科学 財団)	・NSF 助成により得られたすべての重要な結果 (findings) ・NSF 助成により得られる一次データ、サンプル、物理的な収集物、その他作成、収集したサポート材料。 ・NSF 助成により作成されるソフトウェアや発明の共有、それらあるいはそれらにより得られる製品が広く利用できるようにすることを奨励する
DOE (米国エネル ギー省)	・機密扱いに分類されない、制限のない、助成研究成果を立証するのに必要なデジタル・データ。 ・デジタル研究データの定義: 実験、観測、シミュレーションデータ; コード、ソフトウェア、アルゴリズム; テキスト; 数字情報; 画像; ビデオ; 音声; 関係するメタデータ。ローデータ、処理されたデータ、分析データ、公開データ、アーカイブデータなど多様な形式の情報を含む。 *データの収集が実際に行われていない要素については、連邦諮問委員会や公示等の方法により、DOE としてどのようなデータを収集するのが適しているかを判断する。
BBSRC (英国バイオテ クノロジー・生 物科学研究会 議)	・データが論文で使用されているか否かに関わらず助成研究から発生するすべてのデータ ・特に重要と考えられる分野: ①多量的実験から発生するデータ ②長期的なまたは累積的アプローチにより発生する低情報量データ ③システムアプローチから発生するモデル
EC HORIZON2020	研究データの定義: 「研究データ」は、検討・考察のために収集され、推論、議論、または計算のベースとなる、特に事実や数字を意味する。研究のコンテキストでは、データの例として統計、実験結果、測定値、フィールドワークからの観察、調査結果、インタビューの録音や画像が挙げられる。焦点はデジタル形式で提供されている研究データである。 Open Research Data Pilot (※) で適用されるデータ (1) 論文公開時の結果を検証するのに必要なデータ(メタデータを含む)。 (2) “データ管理計画”に記載するその他データ(メタデータを含む)。

※: Open Research Data Pilot とは、プロジェクトによって生成される研究データへのアクセスと再利用の向上と最大化を目的とした取り組み

2. 海外出版社等におけるデータサイテーション、データパブリケーション

海外の主要な出版社からは、オープンなデータジャーナルが相次いで創刊されている。また、学術論文誌に投稿された論文の根拠として基礎データの提出を求める動きや、論文へのデータソースの記述を求めるデータサイテーションの動きが進んでいる。

主要な各社の取り組み状況を見てみると、Nature Publishing Group では、2014年5月にオープンな査読付きデータジャーナル“Scientific Data¹¹”を創刊し、Elsevier 社も、2014年9月、新しいデータ・ジャーナル“Data in Brief¹²”を創刊している。

Public Library of Science 社 (PLOS) では、2014年3月より論文の根拠となる基礎データを、論文公開後すぐに制限なく公開することとした新しいデータ共有方針を、投稿するすべての論文に対し適用した。また、著者に対してはデータセットの再利用について記述される“Data Availability Statement”の提出を要求し、論文冒頭で公開するとしている。¹³

Thomson Reuters 社では、研究データに着目した“Data Citation Index (DCI)¹⁴”をリリースし Web of Science 上で提供を開始した(2012)。DCI では、選定されたリポジトリ(Dryad、Figshare 等)に収録されている研究データの一括検索を可能とし、研究データ引用の促進を目指している。

こうした出版社等における動きは、研究データを学術論文と同じように出版(Publish)し、それが引用(Citation)され、更にはその影響度(Impact)を図るというように、研究データを対象とした新たなビジネスモデル、データサイテーションによる新たな業績評価手法が形成されつつあると言える。

¹¹ Scientific Data <http://www.nature.com/sdata/>

¹² Data in Brief <http://www.journals.elsevier.com/data-in-brief/>

¹³ 2014年02月26日 STIUpdates 学術情報流通ニュース <http://johokanri.jp/stiupdates/policy/2014/02/009675.html>

¹⁴ Data Citation Index http://wokinfo.com/products_tools/multidisciplinary/dci/

3. わが国におけるデータシェアリングの動向

わが国においては、第4期科学技術基本計画(2011年8月19日閣議決定39頁)¹⁵にて、「観測、実験データ等の教育研究成果の電子化による体系的収集、保存やオープンアクセスを推進する」と言及されている。また、2014年12月には、内閣府において「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¹⁶」が発足し、論文のオープンアクセス、データシェアリングを含むオープンサイエンスに関する議論が開始されたところである。

研究コミュニティの側面では、2015年2月、日本学術会議(SCJ)主催で開催された「学術フォーラム 科学研究における健全性の向上—研究活動における不正行為への対応に関するガイドラインへの対応—」においては、「公的な資金によって実施された研究で生み出された成果やその元となるデータ等は公的資産としての性格を有しており、それらを適切に管理・保存し、必要に応じて開示することが、研究者及び研究機関に課せられた責務である」とし、保管すべきデータ等の範囲、必要な措置を定めたガイドラインの作成が求められているとしている。

また日本学術会議では、データ共有に関する世界組織である World Data System(WDS)や Committee on Data for Science and Technology(CODATA)の小委員会が組織化され、情報通信研究機構(NICT)には、2012年に国際プログラムオフィス(IPO)が設置されるなど、国際組織との連携に関する動きがある。

一部の研究分野ではデータシェアリングに関連する取り組みは既に進んでおり、生命科学分野ではJSTのバイオサイエンスデータベースセンター(NBDC)¹⁷、地球環境分野では東京大学地球観測データ統融合連携研究機構が主導するデータ統合・解析システム「DIAS」¹⁸、及び海洋研究開発機構(JAMSTEC)の海洋観測データ¹⁹、物質・材料分野では物質材料研究機構(NIMS)によるマテリアルズインフォマティクスの推進などの事例が挙げられる。

上記のように、わが国においては各ステークホルダーの施策や一部の研究分野ではデータシェアリングの議論も進みつつあるものの、国家としての政策的な議論はまだ始められたばかりの状態である。

諸外国が、オープンデータ、データシェアリング等の手法を駆使した新たな研究プロセス推進に向けた取り組みを進めている状況において、わが国は同様の目的意識を持ちつつも、諸外国と比較して大きく遅れている状況と言える。施策がこのまま進まず世界に大きく水をあけられるということは、地球規模研究への参加機会を喪失し、わが国の国際的ステータスや研究開発力を低下させるとともに、日本の研究業績が世界に正しく評価されなくなる可能性を秘めている。

わが国においては、こうした世界情勢に乗り遅れることなく、研究データを対象とした新たな研究手法による科学技術の発展、国際的な研究開発力強化に向け、効率的かつ効果的な、日本としてあるべきデータシェアリング施策の導入を情報戦略として検討していくことが重要である。

¹⁵ 第4期科学技術基本計画 http://www.mext.go.jp/component/a_menu/science/detail/_icsFiles/fieldfile/2011/08/19/1293746_02.pdf

¹⁶ 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 <http://www8.cao.go.jp/cstp/sonota/openscience/>

¹⁷ バイオサイエンスデータベースセンター <http://biosciencedbc.jp/>

¹⁸ 地球環境情報統融合プログラム <http://www.editoria.u-tokyo.ac.jp/projects/dias/>

¹⁹ 独立行政法人海洋研究開発機構 データベース <http://www.jamstec.go.jp/j/database/ocean.html>