

資料と公共性 : 2018年度研究成果年次報告書

岡崎, 敦

九州大学大学院人文科学研究院 | 九州大学大学院統合新領域学府 : 教授

市澤, 哲

神戸大学大学院人文科学研究科 : 教授

石田, 栄美

九州大学附属図書館 | 九州大学大学院統合新領域学府 : 准教授

後小路, 雅弘

九州大学大学院人文科学研究院 : 教授

他

<https://doi.org/10.15017/2230688>

出版情報 : 2019-03-14. 九州大学大学院人文科学研究院

バージョン :

権利関係 :

研究データ管理の動向及びデータ利活用に向けた課題整理

南山 泰之

1. はじめに

昨今のオープンサイエンスの潮流、および研究公正対応として研究データのオープン化に期待が寄せられている。本稿では、学術コミュニティで用いられるオープンデータと、オープンガバメントの文脈で用いられるオープンデータの認識のずれを踏まえつつ、オープン化の前提となる研究データ管理の動向を紹介する。併せて、研究データの利活用に向けた課題の整理を試みることで、大学がオープンデータの潮流に貢献するための道筋を探る手掛かりを提供する。

2. オープンデータの潮流

今日的な意味での「オープン」は、アクセスのみならず、利用、改変、共有などについても、用途を問わず行うことができるものと定義される¹⁾。その意義としては、

- 1) 品質保証：「オープン」に対する明確な定義を提供することで、「オープン」とそれ以外を区別する
- 2) 互換性：データを組み合わせることが法的・技術的に可能かどうかの指標
- 3) 単純性：データを組み合わせるために複雑な制限（著作権処理等）を要求されない

の3点が挙げられている²⁾。ここで対象とするデータは公共財としての側面を持ち、科学ないし社会を構成する一要素である、と解釈されており、したがって誰が使う場合であっても再利用に制限が生じないことが原則となる。日本においてもこのような理解のもと、官民データ活用推進基本法（平成28年法律第103号）において、国及び地方公共団体はオープンデータに取り組むことが義務付けられている³⁾。

一方、2016年1月に内閣府が策定した第5期科学技術基本計画においては、オープンサイエンスの推進と、国の基本姿勢が述べられている⁴⁾。基本計画によれば、「オープンサイエンスとは、主に論文へのオープンアクセスと研究データのオープン化、すなわちオープンデータを含む概念」とされ、オープンデータが進むことで、研究プロセスの透明化や研究成果の幅広い活用が期待されている。しかしながら、学術コミュニティにおける「オープン」では free access の側面が強調され、free reuse についてはあまり留意されていないことが多いように見受けられる。定義のずれ（あるいは、意識のずれ）が発生している理由はいくつか考えられるが、あり得る一つの説明としては、昨今の「オープンサイエンス」のキーワードのもと、大学図書館／学術出版者、研究者、市民社会といった様々な背景を持つコミュニティが集った結果、各々の文脈に沿った形で「オープン」が解釈されている、と表現できる

だろう。

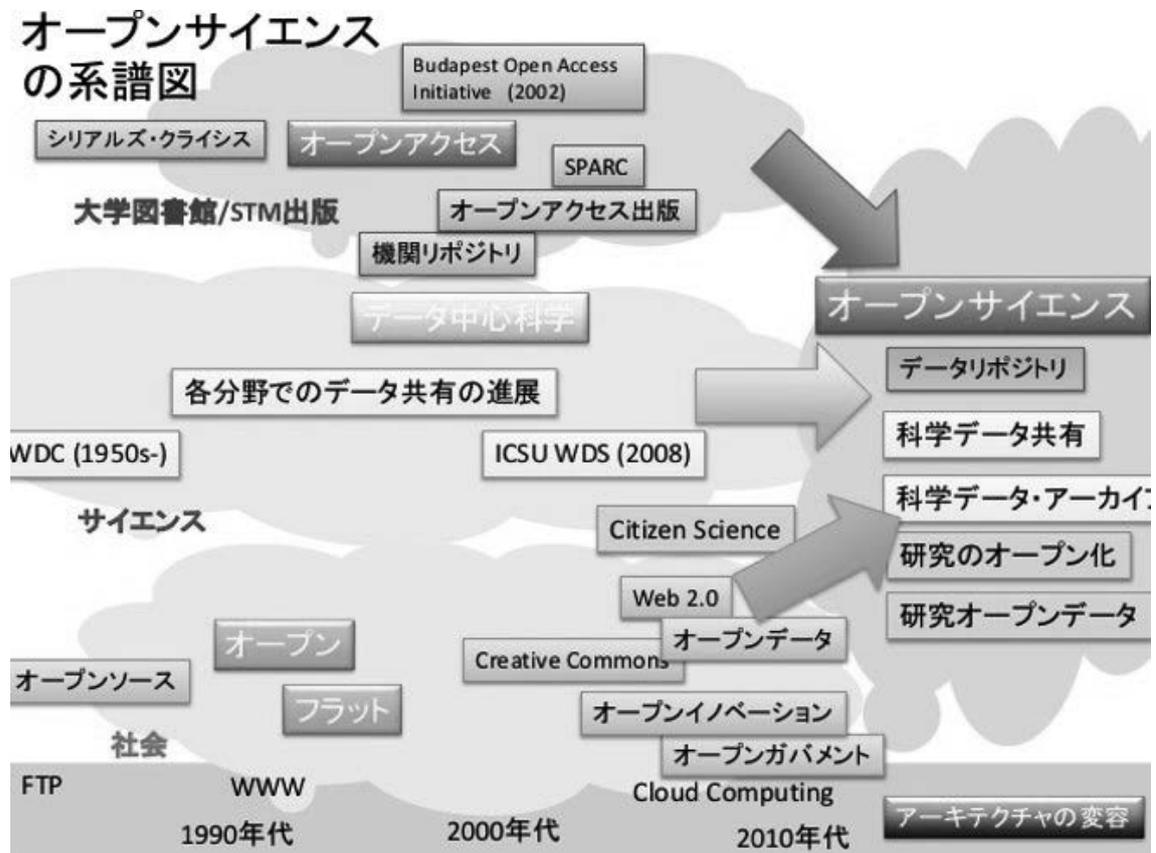


図1 武田英明. オープンサイエンスの系譜図

したがって、一口に「オープンデータ」と言っても、どのデータをオープンにしていくのかによって理解すべき背景と対応が異なる点には留意が必要である。

また、上述のオープンサイエンスとは異なる文脈で、研究不正対応がある。2014年8月に策定された文部科学省のガイドラインから始まる一連の動きにより⁵⁾⁶⁾、社会的な側面からも大学・研究機関は一定期間の研究根拠データの保存を求められており、各大学にてポリシーの整備が進められている⁷⁾。ここで、出自が異なるオープンサイエンス推進と研究公正への対応を考えるにあたり、その手法としての研究データ管理の必要性が分野を超えて俄かに高まりつつある。

3. 研究データ管理の動向

海外における研究データのオープン化については、2013年のG8科学技術大臣会合における合意⁸⁾を重要な契機として、各国で方策が検討・推進されている。また、海外出版者に

よるデータ共有の推奨ないし義務化のポリシー制定も進んでおり⁹⁾、特に論文根拠データの共有、保存が強く謳われている。並行してデータ管理の支援サービスも提供され始めており、サポート窓口の開設、データ取扱いのトレーニング、データ可用性 (availability) のレポート作成¹⁰⁾、機関向けデータ保存ストレージの提供¹¹⁾などが展開されている。大学や学術機関でも大規模大学を中心に事例が蓄積され始めており、研究データ管理ポリシーを定めて指針を示すほか、トレーニングコースの提供、データ管理計画の作成支援、ツールやストレージの提供などのサービスが展開されている¹²⁾。また、コミュニティレベルでもデータ利活用を普及させるための基準策定に係る議論が盛んになされており、特にデータ公開の適切な実施方法を表現している FAIR データ原則¹³⁾、データリポジトリの信頼性を評価するための CoreTrustSeal¹⁴⁾についてはほぼデファクトとなりつつある。

国内においては、研究公正の文脈において 2016 年 2 月に科学技術・学術審議会学術分科会学術情報委員会から審議のまとめ¹⁵⁾が出されており、大学等に期待される役割のひとつとして「技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資する」ことが挙げられている。また、2018 年 6 月には総合イノベーション戦略¹⁶⁾を受けた形で、内閣府の検討会よりデータポリシー策定のためのガイドライン¹⁷⁾が示され、国立研究開発法人は 2020 年度末までにデータポリシーの策定を目指すものとされた。国内の研究資金助成機関においても、2017 年度に科学技術振興機構 (JST) がデータマネジメントプラン (データ管理計画) を要求¹⁸⁾したことを皮切りに、日本医療研究開発機構 (AMED)、経済産業省産業技術環境局などが相次いで独自の計画書提出を要求¹⁹⁾²⁰⁾しており、外部資金を調達する上で体制構築が急務となりつつある。国内最大級の助成事業を持つ日本学術振興会 (JSPS) ではまだデータに関するポリシーを定めていないが、2018 年度より始まった同会の事業²¹⁾では、人文学・社会科学のデータ共有・保存体制構築のため、拠点機関となる 4 機関を中心にデータ整備を進めており、データ専門職の雇用も想定しているとのことで、今後のポリシー策定にも影響を与えることが予想される。

続いて、オープンサイエンス推進の潮流を担う国内コミュニティの活動を紹介する。オープンアクセスリポジトリ推進協会 (JPCOAR) は、2016 年にオープンアクセスを推進する 3 つの団体が集まって結成され、研究者や URA、大学図書館員など様々な職種で構成される団体である²²⁾。運営委員会が設定した重点目標を達成するため、3 つの作業部会と 6 つのタスクフォースに分かれて活動を展開しており、研究データ管理については研究データタスクフォースが中心となり検討を進めている。2017 年に構築、展開された研究データ管理 (RDM) トレーニングツール²³⁾や、同ツールへのフィードバックを元に開発及び試用プロジェクトが進められている新教材²⁴⁾など、主に研究データ管理の普及、教育に重点が置かれた活動が展開されている。また、同時期である 2016 年に立ち上がった研究データ利活用協議会 (RDUF)²⁵⁾では、年に数回、総会やシンポジウムを開催してデータ利活用の知識普及に努めているほか、小委員会の制度を立ち上げ、政策提言や現場への実装を指向した活動を行っている。その他、これらの団体が一堂に会するイベント²⁶⁾も開催されるなど、こちら

も極めて活発である。

一方で、国内大学における研究データ管理体制構築の動きは活発とは言い難い。研究コミュニティに関わりの深いプロジェクトレベルでのデータ管理サービスは散見される²⁷⁾が、大学全体で取り組むことを外向けに明示している事例は極めて少ないようだ²⁸⁾。政策や助成機関によるポリシー制定、コミュニティによる活動の実装が待たれる。

4. データ利活用に向けた課題整理

研究データ管理に関する体制整備が進む中、キープレイヤーとなる研究者自身はデータ共有、公開に対してどのような反応を示しているのだろうか。2018年に実施された国際的な調査によれば²⁹⁾、約76%の研究者がデータの発見可能性を高めることが重要であるとしているものの、データ整理に係る人的労力やコスト、権利関係の不明確さ、利用可能なリポジトリの選択などに問題があると回答している。

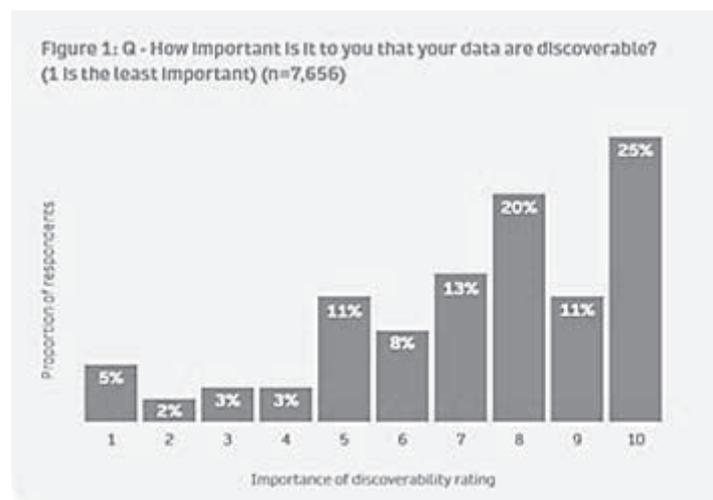


図2 研究者の意識調査

意識調査からも見て取れるように、データそのものの公開には研究アイディアの秘匿、データ公開による研究環境の優位性の放棄といった問題が存在し、また現状では自身の研究推進よりもデータ共有／開を優先させるインセンティブに乏しい。本状況下でオープンデータを推進するに当たっては、オープンデータの理念とは別に、発見可能性を高める工夫、あるいはデータ公開の後押しになる仕掛けが別途必要となる。そこで、本章では、1) 慣習／規則面、2) 情報インフラ、3) 人的基盤の3点から課題整理を試みる。

初めに、慣習／規則面での課題に触れる。研究分野ごとに用いられるデータの種別は大きく異なり、その取扱いもコミュニティに大きく依存する³⁰⁾。データポリシーは機関として

の方針を定めるものであるため、各分野で要求されるデータ管理の実際とはその制定目的が異なる。実効性のある形で整備するためには、社会的に要請される研究公正の側面からのデータ管理だけではなく、学術コミュニティの自律性にも配慮した慎重な調整が必要となる。

続いて、情報インフラの問題に触れる。まず、ソフト面の問題としてデータ管理のシステム基盤をどのような基準に沿って整備するか、が問題となる。利活用の可能性を最大化するためには、データの構造化、ライセンス（データの利用条件提示）、メタデータ付与などを組織のポリシーに沿った形で考える必要がある。また、ハード面においてはデータ管理システム、データバックアップシステム、データ検索システムの整備といった問題が存在し、研究公正を含めた要求を満たすシステム選定は大きな課題となる。ハード面においては現在国立情報学研究所にてこれらを見据えたシステム開発が進められており³¹⁾、対策を検討する上で有力な選択肢となろうが、いずれにせよ現段階において国際的にデファクトといえる基準がある訳ではない。議論に当たっては相応の専門知識や労力を要することもあり、検討のための体制構築が重要なポイントとなろう。

最後に、人的基盤について触れる。昨今のオープンサイエンスの潮流と並行して、データライブラリアン、データキュレーター等のデータ専門職の名称が聞かれるようになりつつあるものの、その役割に対する認知度、普及度は依然として低い。実務的には研究者自身がキュレーターの役割を兼務することも多く、大学／研究機関における人事上の取り扱いは未だ見えないのが現状といえる。もっとも、前述した研究データ管理の教材も含め、データサイエンスやオープンデータをキーワードにした教材は複数存在しており³²⁾、既存の教材を採用の認定基準として用いる、あるいは既存の職種向け能力開発プログラムとして活用する、などいくつかの選択肢は考えられよう。効率的なデータ管理は複数の部署が協力して行う必要があるため、全体像を把握し主体的にサービスを設計できるデータ専門職を人事計画上どう位置付けるべきか、全学的な視点での検討が求められるだろう。

5. おわりに

本発表では、オープンデータを実現するための前段階となる研究データ管理の重要性、国内外における動向及び実現に向けた具体的な課題整理について述べた。フロアも含めたディスカッションでは、データ公開のインセンティブ設計やデータ整備のコスト、データ公開時の責任やオープン化と大学評価との関連性など幅広い論点が展開され、主として社会的側面に対する意識の強さを垣間見ることとなった。データポリシー制定の局面で表面化する課題ということもあり、今後特に比重が大きくなるものと思われる。引き続き議論の深化に期待したい。

- 1) The Open Definition. <http://opendefinition.org/>, (accessed 2019-02-08).
- 2) 東 修作. “なぜオープンデータにとってオープンの定義が重要なのか：品質、互換性そして単純性”. <http://okfn.jp/tag/open-definition/>, (参照 2019-02-08).
- 3) 内閣官房情報通信技術（IT）総合戦略室. “オープンデータ | 政府 CIO ポータル”. <https://cio.go.jp/policy-opendata>, (参照 2019-02-08).
- 4) 内閣府. “第 5 期 科学技術基本計画”. (2016 年 1 月). <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>, (参照 2019-02-08).
- 5) 文部科学省. “研究活動における不正行為への対応等に関するガイドライン”. http://www.mext.go.jp/b_menu/houdou/26/08/_icsFiles/afieldfile/2014/08/26/1351568_02_1.pdf, (参照 2019-02-08).
- 6) 日本学術会議. “（回答）科学研究における健全性の向上について”. <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>, (参照 2019-02-08).
- 7) 例えば、九州大学においては研究データの保存等に関するガイドラインを定め、研究者の責任において研究データの保存・管理を行う旨を定めている。
九州大学. “研究データの保存等に関するガイドライン”. <https://www.kyushu-u.ac.jp/f/1461/guideline.pdf>, (参照 2019-02-08).
- 8) 内閣府. “G8 科学大臣及びアカデミー会長会合の結果概要”. <https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20130620.html>, (参照 2019-02-08).
- 9) 例えば、SpringerNature 社ではデータポリシーを 4 つのタイプに分類し、各タイプのテンプレートを提供している。
SpringerNature. “Research Data Policy Types”. <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096>, (accessed 2019-02-08).
- 10) データサービスの事例として、同じく SpringerNature 社の事例を紹介する：
SpringerNature. “Research data services”. <https://www.springernature.com/gp/open-research/institutions/research-data-services>, (accessed 2019-02-08).
- 11) Figshare. “Figshare now available on custom domains”. https://figshare.com/blog/Figshare_now_available_on_custom_domains/461, (accessed 2019-02-08).
- 12) 例えば、University of Edinburgh. “Information Services : Research Data Service”. <https://www.ed.ac.uk/information-services/research-support/research-data-service>, (accessed 2019-02-08).
- 13) FORCE11. “The FAIR Data Principles”. <https://www.force11.org/group/fairgroup/fairprinciples>, (accessed 2019-02-08).

なお、FAIR データ原則の日本語訳及び成立経緯については以下に詳しい：
NBDC 研究チーム. “データ共有の基準としての FAIR 原則”.

<https://doi.org/10.18908/a.2018041901>, (参照 2019-02-08).

- 14) CoreTrustSeal. <https://www.coretrustseal.org>, (accessed 2019-02-08).
 なお、CoreTrustSeal の概要については以下に詳しい：
 南山泰之. 信頼できるデータリポジトリの中核的な統一要件. カレントアウェアネス-E. 2017, No. 320. <http://current.ndl.go.jp/e1888>, (参照 2019-02-08).
- 15) 文部科学省. “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, (参照 2019-02-08).
- 16) 内閣府. “統合イノベーション戦略”. <https://www8.cao.go.jp/cstp/tougosenryaku/index.html>, (参照 2019-02-08).
- 17) 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”. <http://www8.cao.go.jp/cstp/stsonota/datapolicy/datapolicy.pdf>, (参照 2019-02-08).
- 18) 科学技術振興機構. “オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針”. https://www.jst.go.jp/pr/intro/openscience/policy_openscience.pdf, (参照 2019-02-08).
- 19) 日本医療研究開発機構. “データマネジメントプランの提出の義務化について”. <https://www.amed.go.jp/koubo/datamanagement.html>, (参照 2019-02-08).
- 20) 経済産業省. “委託研究開発におけるデータマネジメントに関する運用ガイドラインを策定しました”. <http://www.meti.go.jp/press/2017/12/20171227001/20171227001.html>, (参照 2019-02-08).
- 21) 日本学術振興会. “人文学・社会科学データインフラストラクチャー構築プログラム”. https://www.jsps.go.jp/data_infrastructure/index.html, (参照 2019-02-08).
- 22) 江川和子. オープンアクセスリポジトリ推進協会の発足. カレントアウェアネス-E. 2016, No. 309. <http://current.ndl.go.jp/e1830>, (参照 2019-02-08).
- 23) 常川真央ほか. 研究データ管理(RDM)トレーニングツールの構築と展開. 情報知識学会誌. 2017, Vol. 27, No. 4, pp. 362-365. https://doi.org/10.2964/jsik_2017_042 (参照 2019-02-08).
- 24) 尾城孝一, 古川雅子. “オンライン講座「研究データ管理サービスの設計と実践」の開発と試用プロジェクト”. https://www.nii.ac.jp/event/upload/libfair2018_forum3_1.pdf, (参照 2019-02-08).
- 25) 研究データ利活用協議会. <https://japanlinkcenter.org/rduf/index.html>, (参照 2019-02-08).
- 26) 伊東敦子ほか. Japan Open Science Summit 2018<報告>. カレントアウェアネス-E. 2018, No. 352. <http://current.ndl.go.jp/e2051>, (参照 2019-02-08).
- 27) 北本朝展ほか. 地球環境情報統融合プログラム DIAS データ共有に基づく社会課題解決. 情報管理. 2015 Vol. 58, No. 6, pp. 413-421.

- 28) 例えば、京都大学情報環境機構. “データ保存サービス”. <http://www.iimc.kyoto-u.ac.jp/ja/services/archive/>, (参照 2019-02-08).
- 29) Stuart, D., Baynes/G., Hrynaszkiewicz/I., Allin/K., Penny/D., Lucraft/M., Astell, /M., Whitepaper. “Practical challenges for researchers in data sharing. 2018”. <https://doi.org/10.6084/m9.figshare.5971387>, (accessed 2019-02-08).
- 30) RDA-CODATA Legal Interoperability Interest Group. (2016, October 20). “Legal Interoperability of Research Data: Principles and Implementation Guidelines”. Zenodo. <http://doi.org/10.5281/zenodo.162241>, (accessed 2019-02-08).
- 31) 山地一禎. オープンサイエンス基盤研究センターの新設について. カレントアウェアネス-E. 2017, No. 327. <http://current.ndl.go.jp/e1925>, (参照 2019-02-08).
- 32) 例えば、九州大学数理・データサイエンス教育研究センター. “講義資料”. (<http://mdsc.kyushu-u.ac.jp/lectures>), “オープンデータ研修ポータル”. (<https://www.opendata-training.org/>) など (参照 2019-02-08).