

CA1858

動向レビュー

データジャーナル：  
研究データ管理の新たな試み

みなみやますゆき  
南山泰之\*

1. はじめに

近年関心の高まる研究データの管理において、データを「出版（パブリケーション）」することでその一翼を担おうとする試みである、「データジャーナル」という新たなタイプの学術雑誌が注目されている。2014年には“Scientific Data”（Nature Publishing Group），“Data in Brief”（Elsevier）の創刊も相次ぎ、データジャーナルへの期待は高まりつつあると言えよう。本稿では、データジャーナルの概要や登場に至る背景を概観しつつ、データジャーナルの主な特徴を解説する。併せて、近年における日本での動向を紹介する。

2. データジャーナルとは

(1) 概要

データジャーナルとは、2014年9月に日本学術会議情報学委員会から出された報告書<sup>(1)</sup>によれば、「データ生産者が分野を超えて連携して、オリジナル論文に埋め込んだデータや論文投稿時に棄却した高品質のデータを学術の成果として集積するための新たな場」と表現される。データリポジトリ<sup>(2)</sup>に保存されている研究データ等について、「学術の成果」として体系的に情報を集約し再利用可能にするために、従来のオリジナル論文と同じ「出版」の枠組みを活用する試み（データパブリッシング）の一形態であり、データから得られた知見（解釈）ではなくデータそのものの解説（データに関する記述。データの保存先情報等も含む）を論文として掲載する点に特徴がある<sup>(3)</sup>。近年のオープンアクセスに係る潮流を反映した形で、多くのデータジャーナルはオープンアクセス誌<sup>(4)</sup>として刊行されており<sup>(5)</sup>、分野を超えた幅広い層によるデータの利活用が期待される。

(2) 登場までの背景

科学的な根拠を担保するため、学術研究の成果は公開され、第三者による検証を経て再現性が認められなければならない。第三者が検証を行うためには、論文のみが公開されているだけでは不十分であり、研究の過程で生じた成果物（研究データ、ソースコード、プ

ロトコル等）も公開されて初めて十分な検証が可能となる<sup>(6)</sup>。

上記の理解のもと、研究成果たる論文の根拠となる研究データ等については、論文の補足情報として、あるいは論文内の図表という形で、検証に必要な範囲のデータが公開されてきた。一方、必ずしも論文に直結しない、より広範なデータ等の公開や共有に関する取り組みは、データ取得コストの高い生命科学や地球科学・天文学、材料科学などの分野に留まっていた<sup>(7)(8)</sup>。しかし、研究助成団体によるデータ管理計画作成の義務化や、オープンサイエンスの推進、あるいは研究不正対策といった国際的潮流・社会的要請<sup>(9)</sup>を受け、海外の大学図書館などではデータリポジトリを立ち上げ、研究データ一般を体系的に保存する取り組みが始まっている<sup>(10)(11)</sup>。

もともと、データリポジトリ運用にも課題はある。収録対象データは運営元のポリシーによって異なるが、一般に玉石混交であると言われており<sup>(12)</sup>、各データリポジトリを横断する検索システムも現段階で充実しているとは言い難い。また、データベースの作成にはメタデータ記述作業が不可欠であるが、手間がかかる上に記述できる情報量に限界があり、さらには多くの場合業績として認められないため、研究者にとっては積極的な動機づけに欠ける<sup>(13)</sup>。このような問題意識のもと、研究データ等の管理について、より効率的な保存や公開、再利用促進のための枠組みが模索されてきた（図1）。

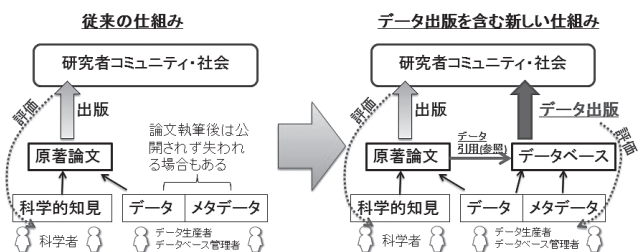


図1 データの利用・評価体制の変革

出典：地球電磁気・地球惑星圏学会，“地球電磁気学・地球惑星圏科学の現状と将来”。2013, 136p.

[http://www.sgepss.org/sgepss/shorai/SGEPSS\\_syorai\\_Jan2013.pdf](http://www.sgepss.org/sgepss/shorai/SGEPSS_syorai_Jan2013.pdf) (参照 2015-08-20).

(3) データジャーナルの登場

研究データに関する情報をより豊かにし、さらに高品質なデータを産出するために、研究データ等を出版し解説を加えるデータパブリッシングは従来からなされてきた<sup>(14)</sup>。したがって、試み自体はさほど新しいものではないが、データジャーナルにおいては、事業モデルがある程度確立した論文出版事業のノウハウを

\*情報・システム研究機構国立極地研究所情報図書室

活用すること<sup>(15)</sup>で、データの定期刊行及び査読による一定の質の担保を実現している点に特徴がある。

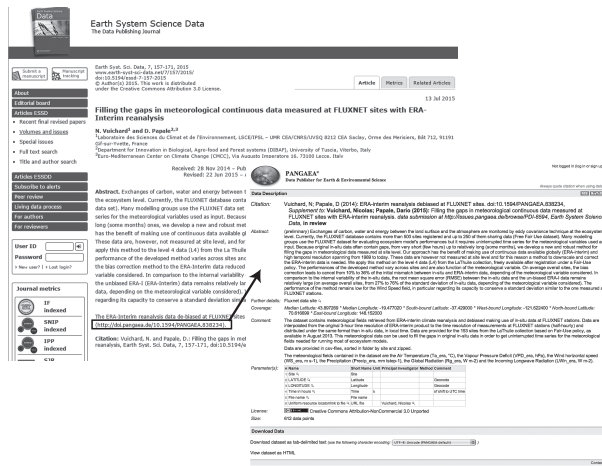


図2 データ論文のスクリーンショット例

出典：http://www.earth-syst-sci-data.net/7/157/2015/essd-7-157-2015.html, (accessed 2015-08-20).

近年ヘルスサイエンス・生命科学、地球科学といった分野を中心にデータ論文の投稿が増えつつある<sup>(16)</sup>のは、データジャーナルへの投稿により後述のようなメリットが期待でき、既存のデータリポジトリの問題点を補完することが可能なこと、また特に地球観測系のデータについては、モニタリングなどの継続的観測により連続したデータが取得され、またそのデータの分析により新たなデータが産出されていく、という研究プロセスがジャーナルという出版形態に合致すること<sup>(17)</sup>、などが注目を集める理由の一端として考えられるだろう。

(4) 投稿のメリット

データジャーナルへの投稿は、すなわち出版の機能である登録、認定、認知（の向上）、保存、評価<sup>(18)</sup>に着目し、研究データの公開や再利用を促進させようとする試みであり、具体的には下記のような効果が期待できる。

a) 研究データ情報の体系的な集約

既存のオリジナル論文からデータの記述を切り離すことにより、従来よりも詳細な情報の掲載が期待できるとともに、独立して再現性の検証が可能となるため、研究の透明性が高まる<sup>(19)</sup>。また、オリジナル論文で引用されず、かつデータベース化が難しい単発のデータについても、データ論文を介してオリジナル論文と共通のプラットフォーム（ScienceDirect等）で段階的・体系的に蓄積しトレースすることが可能になるため、20年間でその80%が失われるとも言われる、学

術の成果としてのデータセット（研究データ、ソースコード、プロトコル等のまとめ）<sup>(20)</sup>の保存と再利用が期待できる。

さらに、データジャーナルは一般に、ある程度特定された分野の主題をその対象としているため、サブジェクトプラットフォームとしての機能も有している。つまり、既存のリポジトリ（figshare<sup>(21)</sup>、Zenodo<sup>(22)</sup>、各大学図書館のデータリポジトリ等）に搭載された研究データ等に関するデータ論文を投稿することで、特定された分野の研究データ情報を集約することが可能になる。

b) 研究データの質の担保

データジャーナルによる一般的な研究データ等の出版プロセスを図3に示す。データセットに対し、著者は①リポジトリへデータセットを保存し、②データセットに関する詳細な記述をデータ論文として投稿する。出版者は③投稿されたデータセット及びデータ論文を確認し、所定の手続きを経た上で④両者に識別子を付与する、という作業を行う。

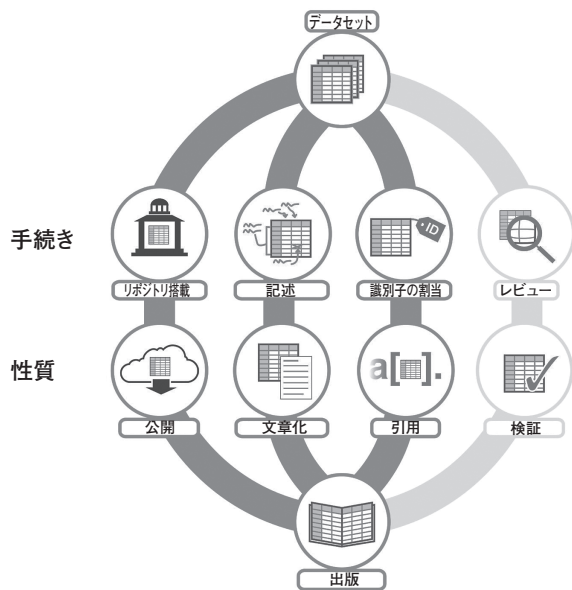


図3 研究データの出版プロセス

出典：Kratz J, Strasser C. Data publication consensus and controversies. F1000Research. 2014, 3, 94. doi:10.12688/f1000research.3979.3.

上述の通り、データジャーナルに投稿されるデータ論文は既存のオリジナル論文と同様に第三者による査読プロセス（③）を経るため、掲載されたデータ論文の対象データについては一定の質が期待できる。もっとも、データ論文の「質」は既存のオリジナル論文の「質」とは意味が異なり、「新規性の高い」論文が評価

されるのではなく、「一定の記述様式（ディスクリプター）に従った、正確かつ豊富な記述がなされている」論文が評価される<sup>(23)</sup>。求められる記述の精粗や重視するポイントはジャーナルごとに異なるが<sup>(24)</sup>、概ね表のような情報が記載されていることが多い。

表 データ論文の記述様式<sup>(25)</sup>

No	項目	説明
1	基本情報	所有者、名前、DOI/URI など
2	利害関係	データセットに関する個人的 / 組織的な関係性
3	範囲	空間的 / 時間的
4	形式	フォーマット、エンコード、言語など
5	ライセンス	—
6	詳細な帰属	各著者ごとの貢献度
7	プロジェクト	—
8	データセットの来歴	取得方法、機材などを含む
9	品質	データの限界値・異常値などの情報
10	再利用	潜在的な価値

c) 流通の促進

研究データの活用・引用に関しては様々な取り組みが既に報告されているが（CA1818 参照）<sup>(26)</sup>、データジャーナルでは出版の枠組みを流用するため、前述の通り既存のオリジナル論文と共通のプラットフォームにデータ論文を掲載することができ、既に確立された検索システムを利用できる。また、研究データのメタデータスキーマはその分野や収録対象となるデータリポジトリによって大きく異なるが<sup>(27)</sup>、データ論文を介することにより、データ論文そのものに付与されたメタデータを利用することが可能になる。すなわち、既存のオリジナル論文と同様のメタデータスキーマによっても研究データの検索が実現できることになる。したがって、データリポジトリ専用の検索システムを構築しなくともよく、また利用者の視点からみると、慣れた検索システムで一元的に検索が可能であることから、視認性の向上、ひいては流通の促進が期待できる。

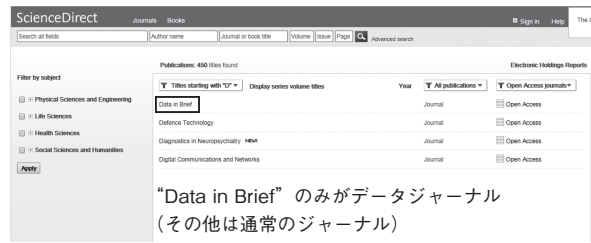


図 4 ScienceDirect 上に並ぶデータジャーナル

出典：<http://www.sciencedirect.com/science/journals/d/open-access>, (accessed 2015-08-20).

d) 研究データの保存やデータ作成者に対するインセンティブ付与

データジャーナルへの投稿を評価の対象とすることで、データに関する記述が業績として認められることになるため、研究者にとってより質の高いデータを作成・保存する動機づけになり得る。また、これまではデータの作成者と論文の著者が異なる場合、研究データは従来まで引用の対象とならなかったため、データの作成者が十分に評価されてこなかった、という問題に対しても、データジャーナルにおいてはデータ作成者が著者となるため、相応の評価を行うことが可能になる<sup>(28)</sup>。

3. 日本の動向

これまで特定の分野内の取り組みに留まっていたデータ・パブリケーションが、日本においてより一般的に言及され始めたのは、管見の限り 2012 年頃からであり<sup>(29)</sup><sup>(30)</sup>、2014 年には前述の日本学術会議による本格的な検討が報告されている。2015 年にはデータジャーナルに関するシンポジウム<sup>(31)</sup>が開催されており、内閣府によるオープンサイエンスに関する検討会における言及<sup>(32)</sup>、文部科学省第 8 期学術情報委員会内での検討<sup>(33)</sup>、国内機関によるデータジャーナル創刊に向けた検討<sup>(34)</sup>など、2014 年末から 2015 年にかけて様々な議論が展開されている。関係各所による今後のさらなる検討とともに、日本発のデータジャーナル創刊といった具体的事例が待たれるところであろう。

4. おわりに

研究データの出版はさほど目新しい取り組みではないものの、学術雑誌の枠組みを利用したデータジャーナルは、昨今の研究データ等の保存や引用の問題を解決する枠組みの一つとして検討に値する。研究データの共有を加速し、技術・実践例等の蓄積を目的とする研究データ同盟 (RDA) (E1531, E1676 参照) でもデータ・パブリケーションに関する Interest Group が立ち上げられ<sup>(35)</sup>、現在も検討が進められているようだ。

また、今後の研究成果の発表は、データジャーナルへ投稿し掲載されたデータ論文をベースにオリジナル論文を執筆する、という二段階を経ることも想定され、研究活動の可視化という視点からは、オープンサイエンスの実現に向けた具体的な試み、という捉え方も可能であるように思われる。研究データ管理のあり方に関する議論の中で、データジャーナルがどのような立ち位置を獲得していくのか、国内外における今後の展開に注目したい。

(1) 日本学術会議情報学委員会国際サイエンスデータ分科会. “オープンデータに関する権利と義務－本格的なデータジャーナルに向けて－”. p. v. <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h140930-3.pdf>. (参照 2015-07-01).

(2) データリポジトリとは、研究データ（画像データや数値データ、メタデータ等）やそれに付随するプログラム等を収集・保管の対象とするリポジトリを指す。データリポジトリに要求される機能、取り組み等に関しては以下に詳しい。南山泰之. 研究データ管理における機関リポジトリの可能性. 大学図書館研究. 2015, vol. 103, in press.

(3) 前掲. p. 12.

(4) データジャーナルにおいても、論文投稿料 (Article Processing Charge: APC) の問題など、主に既存のオープンアクセスジャーナルと同様の問題が起こり得る。なお、この問題に関しては以下に詳しい。三根慎二. オープンアクセスジャーナルの現状. 大学図書館研究. 2007, vol. 80, p. 2-3.

(5) 林和弘, 村山泰啓. 研究データ出版の動向と論文の根拠データの公開促進に向けて: オープンサイエンスをめぐる新しい潮流 (その3). 科学技術動向. 2015, (148), p. 6.

(6) Laine, C., Goodman, S.N., Griswold, M.E., Sox, H.C. Reproducible research: moving toward research the public can really trust. *Annals of Internal Medicine*. 2007, vol. 146, no. 6, p. 450-453.

(7) 高祖歩美. 生命科学分野におけるデータの共有の現状と課題. 情報管理. 2013, vol. 56, no. 5, p. 294-301. doi: 10.1241/johokanri.56.294.

(8) 情報通信研究機構統合データシステム研究開発室. “WDS 関連情報”. <http://www2.nict.go.jp/isd/ISDS-contents/wdsjp.html>. (参照 2015-07-01).

(9) 池内有為. 大学図書館による研究データ管理の最前線: 研究力を強化するエディンバラ大学の事例. 現代の図書館. 2014, vol. 52, no. 4, p. 227-236.

(10) The University of Edinburgh. “Edinburgh DataShare”. <http://datashare.is.ed.ac.uk/>. (accessed 2015-07-01).

(11) Purdue University. “PURR: Purdue University Research Repository”. <https://purrr.purdue.edu/>. (accessed 2015-07-01).

(12) 林和弘, 村山泰啓. 前掲. p. 4-9.

(13) 日本学術会議情報学委員会国際サイエンスデータ分科会. 前掲. p. 7-9.

(14) 例えば, National Institute of Polar Research. “JARE Data Reports”. [http://ci.nii.ac.jp/vol\\_issue/nels/AA00256429\\_en.html](http://ci.nii.ac.jp/vol_issue/nels/AA00256429_en.html). (accessed 2015-08-03). (1968 年創刊)

(15) 林和弘, 村山泰啓. 前掲. p. 6.

(16) Candela, L., Castelli, D., Manghi, P. and Tani, A. Data journals: A survey. *Journal of the Association for Information Science and Technology*. 2015, p. 4 (Table 2). doi: 10.1002/asi.23358.

(17) 村山泰啓. “総括: データジャーナルと観測科学データ統合”. 情報通信研究機構. [http://www2.nict.go.jp/isd/ISDS-contents/study-group/docs/08date\\_journal\\_and\\_observation\\_science.pdf](http://www2.nict.go.jp/isd/ISDS-contents/study-group/docs/08date_journal_and_observation_science.pdf). (参照 2015-07-01).

(18) Van de Sompel, et al. Rethinking scholarly communication. *D-Lib Magazine*. 2004, 10(9), doi:10.1045/september2004-vandesompel.

(19) Welcome, Scientific Data!. *Nature*. 2014, vol. 509, p. 534. doi: 10.1038/509534a.

(20) T. Vines, A. Albert, R. Andrew et al. The availability of research data declines rapidly with article age. *Current*

*biology*. 2014, vol. 24, issue 1, p. 94-97. doi: 10.1016/j.cub.2013.11.014.

(21) Figshare. <http://figshare.com/>. (accessed 2015-07-01).

(22) Zenodo. <https://zenodo.org/>. (accessed 2015-07-01).

(23) 林和弘, 村山泰啓. 前掲. p. 6.

(24) 具体例は、以下に詳しい。国立研究開発法人科学技術振興機構情報企画部. “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 2: 研究データ共有ポリシーに関する調査”. [http://jipsti.jst.go.jp/about/pdf/survey\\_on\\_data\\_sharing\\_policy.pdf](http://jipsti.jst.go.jp/about/pdf/survey_on_data_sharing_policy.pdf). (参照 2015-07-01).

(25) Candela, L., Castelli, D., Manghi, P. and Tani, A. 前掲. p. [8].

(26) Robinson-Garcia, N., Jiménez-Contreras, E. and Torres-Salinas, D. Analyzing data citation practices using the Data Citation Index. *Journal of the Association for Information Science and Technology*. 2015. doi: 10.1002/asi.23529.

(27) Digital Curation Centre. “Disciplinary metadata”. <http://www.dcc.ac.uk/drupal/resources/metadata-standards>. (accessed 2015-08-11).

(28) ヒリナスキエヴィッチ イアン, 新谷洋子. Scientific Data: データの再利用を促進するオープンアクセス・オープンデータジャーナル. 情報管理. 2014, vol. 57, no. 9, p. 633. doi: 10.1241/johokanri.57.629.

(29) 渡邊堯. ICSU 世界データシステム (WDS) について. 学術の動向. 2012, vol. 17, issue 6, p. 11-15.

(30) 村山泰啓. 科学データ・マネージメントと新たなデータ共有・利用. 学術の動向. 2013, vol. 18, issue 9, p. 61-63.

(31) 佐藤正樹. 集会報告 学術フォーラム 「科学を変えるデータジャーナル－科学技術データの共有・利用の新たなプラットフォーム」. 情報管理. 2015, vol. 58, issue 3, p. 228-233. doi: 10.1241/johokanri.58.228.

(32) 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」報告書”. p. 22. <http://www8.cao.go.jp/cstp/sonota/openscience/>. (参照 2015-07-01).

(33) 文部科学省. “第8期学術情報委員会 (第4回) 配布資料. 資料1 学術情報のオープン化の推進について (中間まとめ案)”. [http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/036/attach/1360653.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/attach/1360653.htm). (参照 2015-08-18).

(34) 矢吹裕伯. 国立極地研究所北極域データアーカイブの紹介と今後の取り組み. ジャパンリンクセンター. p. 13. [https://japanlinkcenter.org/top/doc/150227\\_5\\_yabuki.pdf](https://japanlinkcenter.org/top/doc/150227_5_yabuki.pdf). (参照 2015-07-01).

(35) “RDA/WDS Publishing Data IG”. RDA. <https://rd-alliance.org/groups/rdawds-publishing-data-ig.html>. (accessed 2015-07-01).

[受理: 2015-08-20]

Minamiyama Yasuyuki

Data journal: a new experiment in research data management