

オープンサイエンスをめぐる新しい潮流(その3) 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

林 和弘 村山 泰啓

概 要

研究データの管理、保存と共有に関する議論が最近盛んであり、国際的な枠組みにおいても研究データの利活用への積極的な取組の検討が各国で進められているが、実際の施策に落とし込むためには課題が多い。特に、分野ごとにどのレベルの研究データを誰がどのように責任を持って保存するか、あるいは、そのデータの質はどのように保証されるかが課題となっている。

一方、昨今出版者によるデータジャーナルの創刊が始まっており、オープンアクセス論文の出版事業の仕組みを活用しているため、他の手探りで行われている手法に比べ、事業の持続性が高い。

当面、研究成果公開のメディアとして確固たる地位を築いている学術雑誌の論文主張と裏付けるデータに関して、しかるべき公開、保存体制を取ることを一つの軸とした施策を検討すべきである。データ出版体制においては、日本独自のデータ出版の可能性を探る方向と、国際的なデータ出版の枠組みの中で日本の一定のプレゼンスを示す方向の両面を検討する必要がある。データジャーナルの質の保証に関しては図書館の活動に強みを活かせる可能性がある。

また、研究助成団体等、研究資金を提供するセクターは、助成研究の成果を公表する際にはその論拠となるデータの公開を促し、引用・参照が可能となる体制・運用を検討する必要があり、大学・研究機関との連携や研究者の意識啓発活動が必要である。

キーワード：オープンサイエンス，研究データ，データジャーナル，図書館，研究助成団体，オープンアクセス，研究論文

1 はじめに

研究データの管理・保存と共有に関する議論が最近盛んであり、政府レベルではG8やGRC(Global Research Council)を通じて、すでに国際的な枠組みの中において、公的資金で行われた研究データの利活用に積極的に各国取り組んでいる¹⁾。研究データの共有を促し、長期的な保存管理体制を整えることは、科学の発展やイノベーションを促して、科学研究そのものの在り方を変えるだけでなく、産業の革新をも促し人類の生活の質の向上に役立つ。その

一方、実際の施策に落とし込むためには課題がまだ山積している状態でもある²⁾。

本稿では、研究データの管理・保存と共有を進める上で、より具体的な政策に関する議論が可能となるトピックとしてデータ出版の国際動向を解説し、日本が取るべき方策について考察を加える。

2 研究データの公開、保存、管理に関する課題

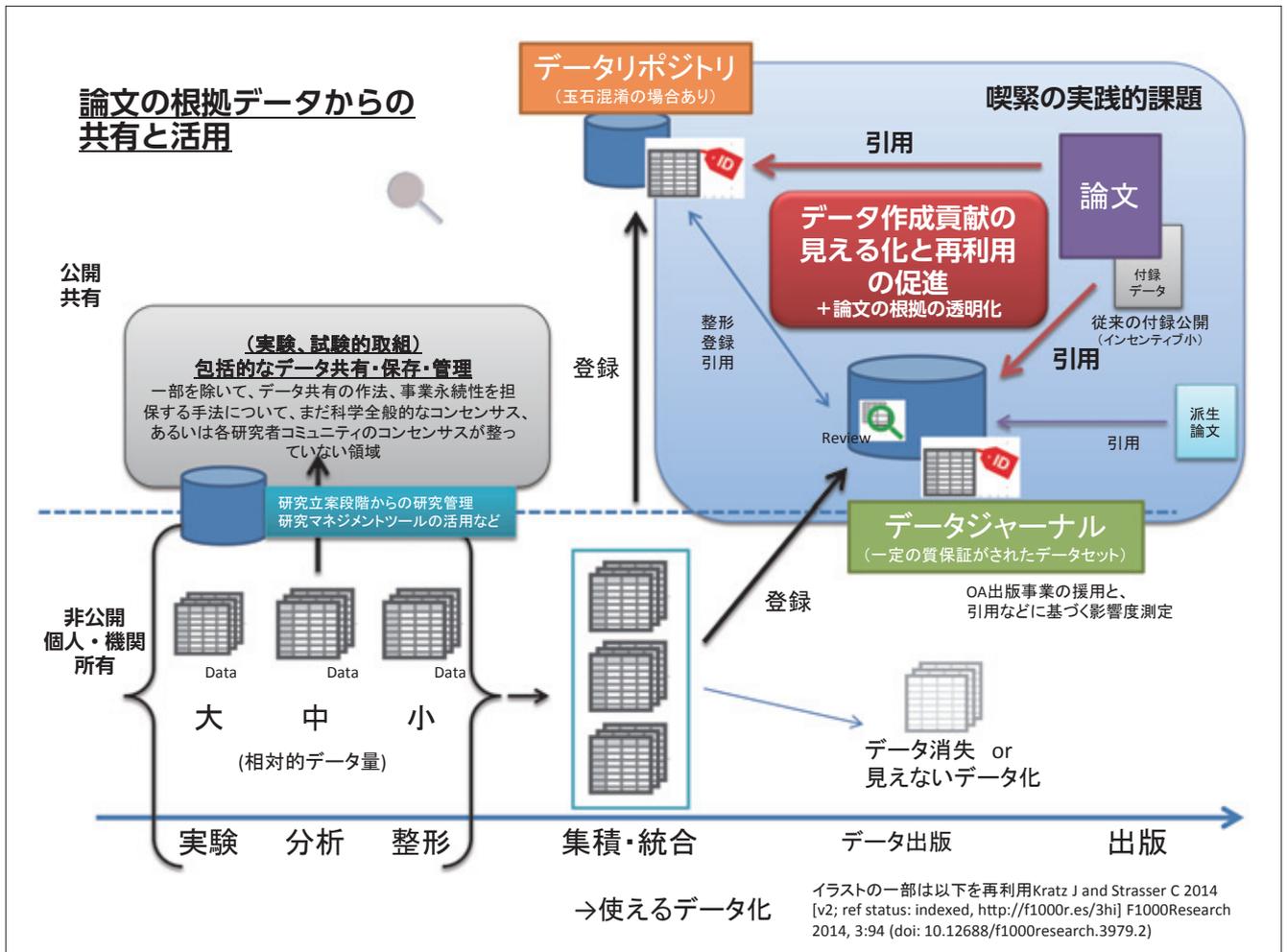
現在、研究データの管理・保存と共有を議論する際に問題となるのは、分野ごとにどのレベルの研究データを誰がどのように責任を持って保存するか、あるいは、そのデータの質はどのように保証されるかである。例えば、図表1左下のように、実験、分析、整形と、それぞれのフェーズにおいて研究データはまとめることができ、研究マネジメントの観点からみれば、原理的には全てのデータを保存すべきではあるが、データ量や手間、経費を考慮すると現実的ではない。遺伝子のGenBank³⁾や結晶のX線構造解析データを集めるケンブリッジ結晶学データセンター（Cambridge Crystallographic Data Centre：CCDC）⁴⁾など、一部の研究分野においては、研究データのデファクトスタンダードのデータ、ないしはデータリポジトリに登録することが慣習として確立されており、他の領域にも同様の動きが見られつつあるが、科学技術全体としてみるとまだこのようなデファクトのデータリポジトリが整っていない

分野の方が多い。

また、各々のデータベース、リポジトリの個々のデータの質の管理もまちまちである。CCDCのように、データチェックの後にデータの質に応じて受け入れを拒否するものもあるが、figshare⁵⁾、DRYAD⁶⁾等、最近構築されたデータリポジトリに関して、データの質のコントロールについては、データ作成者側に委ねられているものが多い。Altmetrics⁷⁾などを利用してデータ公開後のインパクト計量によって質が判断されるという考え方も生まれているが、いずれにせよ、データリポジトリに登載されているデータの質には差があり、利用者の見識が問われるため、専門外の利用には一定のリスクが伴う。

以上、研究データについて分野を問わない広い範囲で管理・保存し、共有を促すための手法として一定のコンセンサスを得ているプロトコルはまだ存在しておらず、また、一部を除いて分野ごとにしかるべきデファクトの手法がくまなく存在している状態でもない。したがって、中長期的な展望の元に、実験的な研究データの管理と保存の試みを繰り返す必要がある。

図表1 データの生成・登録とデータと論文の出版の関係図



3 データ出版とデータジャーナルの動向

3-1 データジャーナルの創刊

昨今、出版者を中心としたデータ出版がデータジャーナルの創刊という形で始まっている。文献8を参考に改訂した最近の主なデータジャーナルを図表2に示す。従来の商業出版者や学会出版者に加えて、最近になって生まれたオープンアクセスを専門とする出版者や、CODATA (Committee on Data for Science and Technology: 科学技術データ委員会) のようにデータの保存と管理に取り組んできた団体によるものもある。

これらのデータジャーナルの多くはオープンアクセス論文を公開する仕組みを援用し、著者側が支払う Article Processing Charge (APC: 掲載料) を利用して出版事業を成り立たせている⁹⁾。いわゆる情報の発信側に課金して、オープンに公開するスタイルであるために、データへのアクセスは誰でも可能となる。また、ほとんどの場合、データの再利用・改変を含むさまざまな利活用が可能となっている。事業モデルがある程度確立したオープンアクセス論文出版事業のノウハウを活用できるため、他の手探りで行われている手法に比べて持続性がある。最近では CrossRef 等の出版者主導の団体が DataCite 等の新しいイニシアチブと積極的に連携を行い¹⁰⁾、

論文と引用/被引用の関係と同じ構図をデータ出版においても積極的に整備しようとしている。

また、もともと雑誌によっては論文には論文をサポートする、あるいは、誌面に掲載しきれない情報を付録 (Supplemental Materials, Supplemental Information) として掲載している。しかし、この場合のデータはあくまで「付録」であって、昨今のデータの公開による研究者の貢献をより積極的に認めようとする動きにはそぐわない。データジャーナルの創刊によって、データセットなどの研究データが「出版物」として識別子 (ID) と共に明示的に公開され、引用が可能となることで、研究者の貢献の見え方に新しい可能性を与えることとなった。

3-2 データの質をコントロールするデータ記述様式 (データディスクリプター)

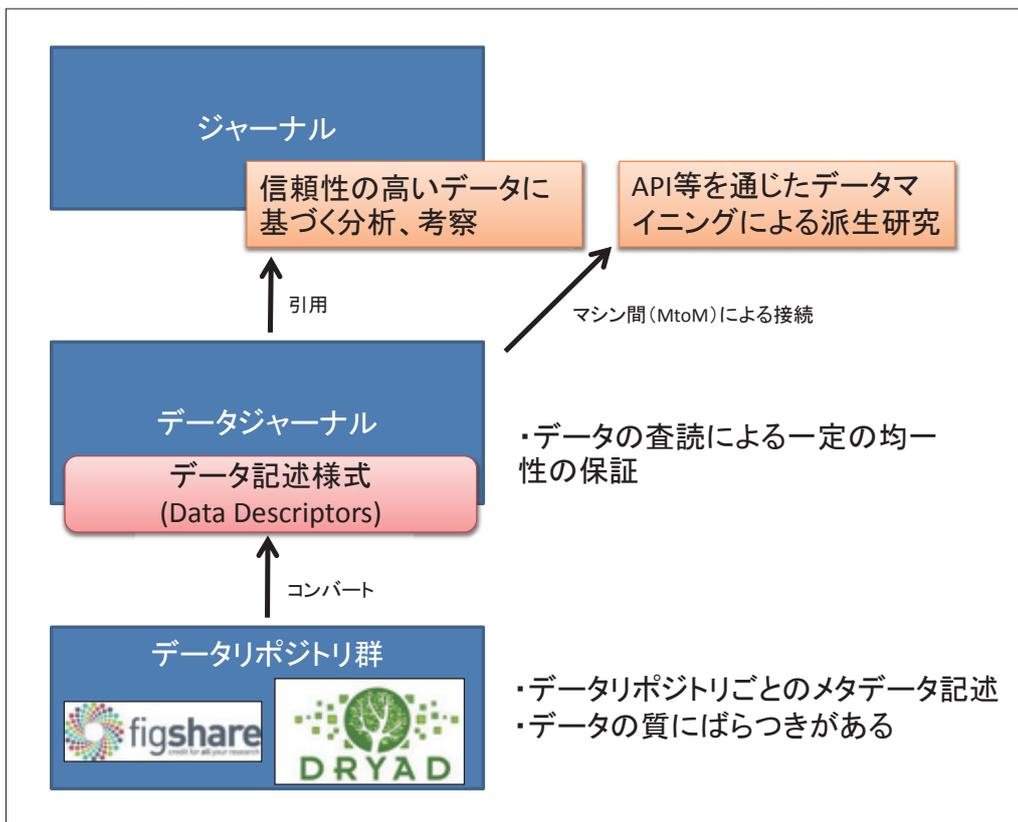
データジャーナルの質の保証は、通常の研究論文とは異なる性質を持つ。データの質保証においては、データがどのように作られたかの素性や使い方についての論説・記述、および長期保管を前提に後世の人間やデータ処理のための機械が判読できる形式でデータを記述すること (データディスクリプター) が重要であり、このデータディスクリプターがあることで、データの内容の継承、データの種類の問わない均一な検索、関連出版物とのリンク付けや、データマイニングが可能となる (図表3)。例

図表2 主なデータジャーナル

No.	ジャーナル名	出版者	タイプ	OA*	URL
1	Geoscience Data Journal	Wiley	商業出版	y	http://www.geosciencedata.com
2	Scientific Data	Nature	商業出版	y	http://www.nature.com/sdata/
3	Data in Brief	Elsevier	商業出版	y	http://www.journals.elsevier.com/data-in-brief
4	International Journal of Robotics Research Data Papers	Sage	商業出版	n	http://ijr.sagepub.com/
5	Biodiversity Data Journal	Pensoft	商業出版	y	http://www.pensoft.net/journals/bdj/
6	BMC Research Notes	BMC	新興OA出版	y	http://www.biomedcentral.com/bmcresnotes/
7	Dataset Papers in Science	Hindawi publishing	新興OA出版	y	http://www.datasets.com/
8	Earth System Science Data	Copernicus	新興OA出版	y	http://earth-system-science-data.net/
9	Ubiquity Press metajournals	Ubiquity Press	新興OA出版	y	http://www.metajnl.com/
10	F1000 Research	F1000 Research	新興OA出版	y	http://f1000research.com
11	GigaScience	BioMed Centralと中国のBGI (旧・北京ゲノム研究所)	新興OA出版	y	http://www.gigasciencejournal.com/
12	Ecological Archives - Data Papers	ESA (Ecological Society of America)	学会出版	y	http://esapubs.org/archive/
13	The Journal of Chemical & Engineering Data	American Chemical Society	学会出版	n	http://pubs.acs.org/journal/jceaaax
14	Journal of Physical and Chemical Research Data	American Institute of Physics	学会出版	n	http://jpcrd.aip.org/resource/1/jpcrbu
15	CODATA's Data Science Journal	CODATA	学会出版	y	http://www.codata.org/publications/data-science-journal

*OA オープンアクセスかどうか(y/n)

図表3 ジャーナル、データジャーナル、データリポジトリの関係図



例えば、NPG社（Nature Publishing Group）では、Scientific Data誌において、データディスクリプターをデータジャーナルの根幹要素と位置づけ、先に述べたデータリポジトリとの違いを明示している¹¹⁾。データディスクリプターを持つデータジャーナルはいわば、データのゲートキーパーとして、ジャーナルや、後世を含む読者との情報の均一的な接続を請け負うこととなる。

4 現実的な研究データの管理・保存と共有に関する喫緊の提言

4-1 研究論文に付随するデータの共有と保存

本格的な研究データマネジメントや、長期的な研究データの管理と保存体制が整い、事業性、持続性が担保されるまでには、現時点ではさまざまな課題が多く解決までに時間を要する。したがって、当面、研究成果のメディアとして確固たる地位を築いている論文の主張に付随するデータに関して、しかるべき公開と共有および保存の体制を取ることを一つの軸とした実践的な施策を検討することが必要である。このことは、論文の主張に

対する、遡及性、再現性を担保することにもなり、研究のコンプライアンスや倫理の観点からも重要な観点である。

4-2 データジャーナルの刊行とライブラリアンの新たな役割

データジャーナルの刊行は、一つのトレンドとして注目し、日本においても学協会を中心とした検討が望まれる。データ出版体制においては、学協会や既存のNII（National Institute of Informatics：国立情報学研究所）やJST（Japan Science and Technology Agency：科学技術振興機構）など情報流通事業体を中心とした日本独自のデータ出版の可能性を探る方法と、分野ごとの国際的なデータ出版の枠組みや学協会の国際連携の中で、例えば、日本の研究者がデータジャーナルの編集や取り決めに関する委員に加わるなど一定のプレゼンスを示す方法を考慮する必要がある。日本学術会議においても情報学委員会国際サイエンスデータ分科会を中心にデータ出版の重要性が説かれている¹²⁾。より具体的な方策として、日本学術振興会の科学研究補助金の成果公開促進費にデータジャーナル創刊の категорияを設ける、ないしは、現行のcategoryの範疇で後押しすることも検討に値する。

また、データジャーナルの質の保証において重要な、データディスクリプターの質のコントロールは、図書館のメタデータ管理との親和性が高いため、例えば、目録管理に代表される機関リポジトリからの研究データ発信の際に、その強みを活かせる可能性がある。研究データ同盟 (Research Data Alliance: RDA) においても、「研究データのための図書館」が一つのテーマ (Interest Group) として討議が行われている¹³⁾。NII では JAIRO Cloud を通じて多大学の機関リポジトリをクラウド上で集約化させているが、これに、さらに、分野ごとのデータディスクリプターを整備し、大学等機関リポジトリからのデータ出版機能を包括的に付加・強化することも一例として考慮に値する。

4-3 研究助成団体の立ち位置と大学等との連携

研究助成団体等、研究資金を提供するセクターは、日本の公的資金を得た研究に対して成果を公表する際にはその論拠となるデータを公開し、引用・参照が可能となることを促し、確認する体制・運用を研究者の負担やインセンティブを考慮しながら整えていくことになる。効率を考慮すれば、研究論文とその付随データについてまず検討する必要がある。その場合、データジャーナルや分野のデファクトとなるデータリポジトリとの連携を意識することになるが、研究助成団体自身がデータリポジトリを構築し、まだデータジャーナルやデファクトデータリポジトリが存在していない分野の研究論文のデータを保存することも考える必要がある。

また、研究マネジメントの観点から見れば、大学等の研究機関においても所属の研究者が出版した論文の論拠データに関して無関心では居られないため、大学・研究機関および、リサーチ・アドミニストレーターとの能動的な連携も考えられる。

4-4 継続的な研究者への啓発活動の必要性

研究論文の付随データといえども、研究者の共有の意識が薄いことが分かっている。例えば、オープンアクセスジャーナルの一つ PLOS ONE では、著者に対して、論文の根拠となったデータの公開を義務づけているが、最近行なわれた PLOS ONE の著者に対する調査¹⁴⁾でも、データの公開を実際に行っている率が低く、義務であることに気づいていないことや、共有に対して消極的な研究者が一定の割合でいることが分かっている。また、他の研究者のデータは使い易ければ利用したいが、自分のデータは利用可能になっていない、という調査結果もある¹⁵⁾。研究者は、少なくとも公的資金を得た研究に対する論文とその根拠となるデータを共有することは、その研究者の貢献が正当にかつ透明性高く認められるための必須の作法であることを認識する必要がある、関係者とのコンセンサス作りが必要となる。例えば、NII の SPARC Japan では、オープンアクセスに関する啓発活動を 10 年以上にわたって行なっており¹⁶⁾、このような活動を拡張する必要がある。

5 おわりに

本稿の提言は、研究データの共有と利活用に関してより具体的な施策につなげるための短期的視点による提言であり、将来的には研究実行段階から適切なデータ管理体制が整うことで、より多様なデータの共有と利活用が進むことが期待される。中長期的視点の試み、および国際的動向については次号以降の本誌にて改めて解説することとしたい。

参考文献

- 1) 村山泰啓, 林和弘. オープンサイエンスをめぐる新しい潮流 (その 1) 科学技術・学術情報共有の枠組みの国際動向と研究のオープンデータ. 科学技術動向. 2014, 146, p.12-17 : <http://hdl.handle.net/11035/2972>
- 2) 村山泰啓, 林和弘. オープンサイエンスをめぐる新しい潮流 (その 2) オープンデータのためのデータ保存・管理体制. 科学技術動向. 2014, 147, p.16-22 : <http://hdl.handle.net/11035/2990>
- 3) GenBank : <http://www.ncbi.nlm.nih.gov/genbank/>
- 4) ケンブリッジ結晶学データセンター : <http://www.ccdc.cam.ac.uk/>
- 5) figshare : <http://figshare.com/>

- 6) DRYAD : <http://datadryad.org/>
- 7) 林和弘. 研究論文の影響度を測定する新しい動き—論文単位で即時かつ多面的な測定を可能とする Altmetrics—. 科学技術動向. 2013, 134, p.20-29 : <http://hdl.handle.net/11035/2357>
- 8) A list of Data Journals (in no particular order) : <http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList>
- 9) 林和弘. オープンアクセスを踏まえた研究論文の受発信コストを議論する体制作りに向けて. 科学技術動向. 2014, 145, p.19-25 : <http://hdl.handle.net/11035/2964>
- 10) CrossRef and DataCite announce new initiative to accelerate the adoption of DOIs for data publication and citation : <https://www.datacite.org/CrossRefDataCiteinitiative>
- 11) 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, 57 (9), p. 629-640. : <http://dx.doi.org/10.1241/johokanri.57.629>
- 12) 報告 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—. 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 : <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h140930-3.pdf>
- 13) RDA Libraries for Research Data : <https://rd-alliance.org/group/libraries-research-data.html>
- 14) Richard Van Noorden. Confusion over publisher's pioneering open-data rules. Nature. 2014, 515, p. 478. doi:10.1038/515478a
- 15) Richard Monastersky . Publishing frontiers: The library reboot. Nature. 2013, 495, p. 430-432. doi:10.1038/495430a
- 16) 国際学術情報流通基盤整備事業 (SPARC Japan) : <http://www.nii.ac.jp/sparc/>

..... **執筆者プロフィール**



林 和弘

科学技術動向研究センター 上席研究官

専門は学術情報流通。1990年代後半より日本化学会英文誌の電子化と事業化に取り組み、オープンアクセスにも対応した。電子ジャーナルから発展する研究者コミュニケーションの将来と、学会、図書館、大学の変革およびオープンサイエンスに興味を持つ。



村山 泰啓

科学技術動向研究センター 客員研究官

専門は超高層大気物理学・リモートセンシング。アラスカでの成層圏・中間圏観測に長く携わり、実験観測データベースの開発も行ってきた。ICSU-WDS 科学委員会 ex officio 委員、国立極地研究所南極観測審議委員、京都大学生存圏研究所客員教授、日本地球惑星科学連合・理事などを歴任。