

# 学術における永続的識別子としての DOI の これまでと今後について

武田 英明\*

キーワード：DOI, 永続的識別子, オープンサイエンス



本稿は、クリエイティブ・コモンズ表示 4.0 国際 (CC BY 4.0) ライセンスの下に提供する (<http://creativecommons.org/licenses/by/4.0/>)。

## 1. はじめに

本連載では DOI をはじめとする情報の流通に関する識別子が紹介されてきた。本稿ではその中で特に学術情報流通という視点から、DOI を中心とした永続的識別子の役割や発展の経緯を振り返り、今後を展望する。

## 2. DOI はなぜ成功したか

DOI は 1987 年の国際 DOI 財団の発足から 20 年以上にわたって維持され、これまで 2.4 億件以上の識別子が発行されてきた。現在も月あたり 5 万件程度が新たに発行され、月あたり 5 億回程度の解決 (resolution ; DOI 識別子が参照され、URL を返答すること) が行われている。これは十分に成功した識別子システムであるといえよう。

識別子が成功するには技術や標準といった問題だけでなく、本連載第 1 稿で宮澤氏が述べたように運用面も大きく関わっている<sup>1)</sup>。その点を鑑みて、DOI の成功の要因を 3 点あげることができる。

### (1) オープンかつ透明性のあるシステム

DOI を運用するシステムの根幹はハンドルシステム (Handle System) である。ハンドルシステムはインターネット上に存在するデジタルオブジェクトなどの資源に永続的識別子を付加管理する技術仕様であり、これに基づき CNRI (Corporation for National Research Initiatives) がシステムを開発し、公開している。DOI もこのシステムを利用している。さらにその上で運用される DOI そのものも ISO において規格化されている。

このように DOI は非プロプライエタリシステムとすることで、透明性や継続性が確保され、公共的なサービスとしての信頼を得やすくなっていると言える。実際、Registration Agency (RA) として新規参入するときも公開されたソフトウェアを利用して容易に開始することがで

きる。

### (2) Crossref の成功

DOI のこれまでの発展の多くは Crossref の貢献によるところが大きい。Crossref は電子化された学術論文へのアクセスを確保することが目的として設立され、技術的にはハンドルシステムに基づく DOI を採用した。Crossref (法人としては Publishers International Linking Association, Inc. (PILA)) はそもそも大手出版社と大手学協会との強力な後ろ盾で設立したものであるが、設立後は NPO として中立的に運営されてきた。その結果、大手から中小まで広く学術出版社・学協会が参加して、学術文献には DOI が事実上必須であるという世界を作り上げた<sup>2)</sup>。識別子が価値を持つには十分にコンテンツがカバーされていることが条件であるが、Crossref は運営においてその目標を達成している。

### (3) 分散管理の成功

DOI は DOI 財団 (The International DOI Foundation) が管理するが、実際の DOI の発行業務は Registration Agency (RA) が委託を受けて行う。ISBN も同様の階層構造をもっているが、DOI は ISBN と異なり、RA の所管が排他的ではない。このため、RA はある意味で競争的である。Crossref は学術文献の DOI において圧倒的なシェアをもつが、DataCite や JaLC<sup>3)</sup> も同様に学術文献の DOI を発行している。既存の RA が特定のドメインを支配することはないため、既存の RA も常に改善が必要であるし、不活発な RA によるドメインの不必要な支配もない。この結果、サービスの改善や新規のドメインを対象とする RA の参入などがおきている。

## 3. DOI の今後

DOI は過去 20 年において成功している。では今後、どのような展開が見込まれるだろうか。学術分野における DOI の今後の展開について述べる。

### 3.1 対象の拡大

Crossref はジャーナル論文への DOI から始まり、学術書籍、モノグラフ等へ対象を拡大してきた。Crossref の対象は基本的には出版社が発行する学術文献である。一

\*たけだ ひであき 国立情報学研究所

〒101-8430 千代田区一ツ橋 2-1-2

E-mail: [takeda@nii.ac.jp](mailto:takeda@nii.ac.jp)

<https://orcid.org/0000-0002-2909-7163>

(原稿受領 2021.2.15)

方、DataCiteは学術機関が公開する研究データを主たる対象として始まった。DataCiteは狭義の研究データに限らず、born digitalなものも積極的に取り込んでおり、これを利用するfigshare<sup>注2)</sup>では画像やスライドなど論文にかぎらない様々なデジタルコンテンツにもDOIを発行している。JaLCも初期はジャーナル論文が主であったが、研究業績すべてにDOIを発行することを目標に、研究データ等にDOIを発行している<sup>注3)</sup>。特にJaLCでは大学等が運営する機関リポジトリでの公開物にDOIを振る仕組みがあるため、大学の各種の成果物にDOIを付与することを可能としている。

この傾向はオープンサイエンスの考え方の普及と相まって、さらに加速することが予想される。オープンサイエンスにおいては、最終結果としての論文や研究データといった研究の成果物だけが公開されるのでは不十分であるとされる。研究の評価や公正性、再利用性といった諸側面においては研究の過程での情報の公開が必要である。今後は、こうした研究過程での情報にもDOIが振られていくことになると思われる。

### 3.2 メタデータとのより強い連携

DOIの基本的機能はDOI識別子(DOIネーム、https://doi.org/で始まるURL)をコンテンツにアクセスできるURLに変換するというものである。現在はこのサービスに必要な最小限のデータ(メタデータ)のみをハンドルシステムで保持している<sup>注4)</sup>。一方、RAはそれぞれ自身のカバーするコンテンツに適したメタデータデータベースを構築して、このメタデータを提供するサービスを提供している。

識別子はメタデータとセットとなることで実用的価値がある。しかし現状のDOIにおいて、識別子とメタデータの紐付けはRA依存になっている点が課題である。

メタデータの相互運用性はDOIの価値を高めるに重要な要素である。特に学術に関するDOIは複数のRA(Crossref, DataCite, JaLC, mEDRA, KISTI, CNKI, ISTIC, Airiti)が発行している。一つの方法で標準的なメタデータを獲得できることが望ましい。このための方策が現在、いくつか行われている。

DOI Citation Formatter<sup>注5)</sup>では、DOIネームから複数の学術関係RAが発行するDOIのメタデータ(書誌情報)を希望する書誌フォーマットで出力してくれるサービスである。現在、前述の学術関係RAのうちCNKIを除く7つのRAがこのサービスに参加している。図1に利用例を示す。このサービスの実装においては、httpプロトコルのredirect機能を使い、個別のDOIネームへの問い合わせをDOIハンドルシステムへの問い合わせではなく各RAのデータベースへの問い合わせに変換するということが行われている。各RAのデータベースが本サービスで規定した簡易的なメタデータを返答することで、統一的な振舞いを実現している。なお、このサービスを使うのではなく、返答の表現を指定したhttp requestをhttps://doi.org/に投げることで直接、構造化されたメタデータを得ることもできる。図2は図1と同じDOIに関する返答である。

### 3.3 DOI間の関係

DOIは他のDOIと関係をもつ。学術DOIであれば、もともと重要な関係は論文の被引用関係である。CrossrefやJaLCでは引用情報をメタデータとして保持している。こ

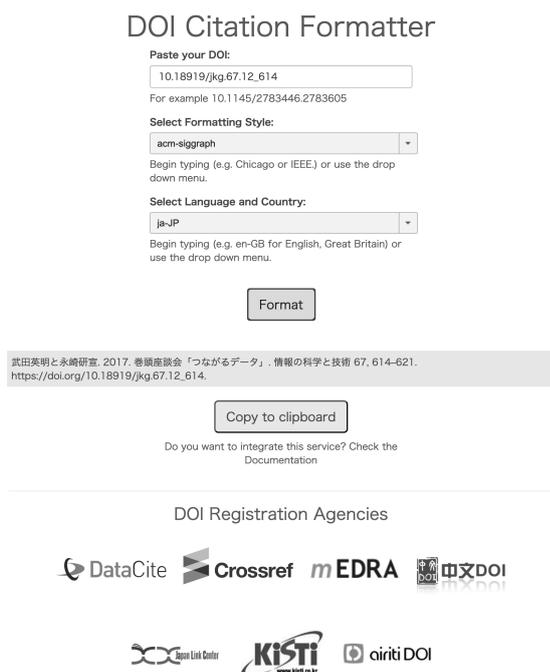


図1 DOI Citation Formatterの利用例

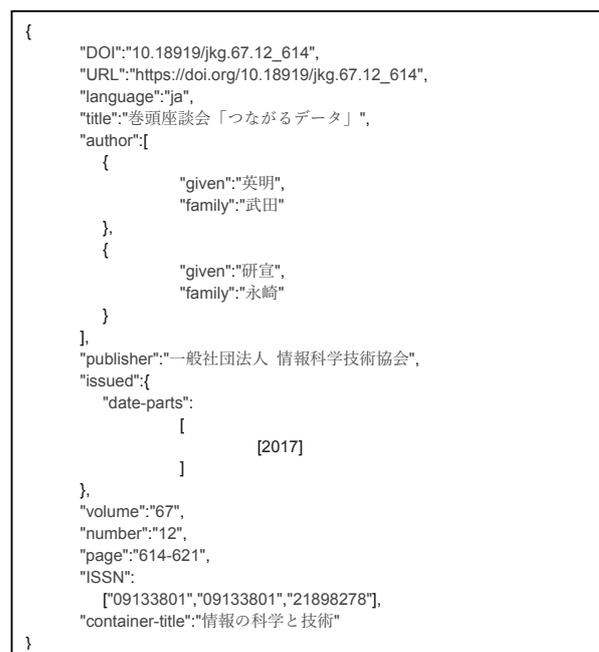


図2 DOIネームに対するメタデータの問い合わせの返答例

れまでは Crossref に登録している出版社の多くでは引用情報を公開してこなかったが、OpenCitation の活動を通じて公開されるようになってきた。

OpenCitation が公開する COCI データセットにはおよそ 6 千万個の Crossref の DOI 間の 7.6 億の引用関係が含まれている (2020 年 12 月)<sup>注6)</sup>。これにより巨大な DOI のネットワークが形成されている。

また、DataCite では全体部分関係やバージョン関係など、DOI 等の識別子との多様な関係づけを行えるメタデータを用意している。

### 3.4 DOI と他の PID との連携

学術においては本連載で紹介されたように様々な識別子が使われるようになってきている。識別子は相互につながり合うことで自身の信頼性と有用性を増すことができるので、識別子間の連携は重要である。現在、多様な識別子の連携が模索されている。以下ではそのいくつかを紹介する。

Crossref, DataCite, JaLC 等ではすでに ORCID ID をメタデータの著者情報として含めることができるようになっており、DOI のメタデータとして ORCID ID 連携が実現している。Crossref では他にも論文の補助金受給先を記述するための Open Funder Registry<sup>注7)</sup> の識別子<sup>注8)</sup> をメタデータに埋め込むことも可能にしている。

DataCite Commons<sup>注9)</sup> というプロジェクトにおいては、Crossref や DataCite 等の書誌データにおける論文 (DOI)、著者 (ORCID)、組織 (ROR) の関係を PID Graph として集積して、検索サービスとして利用可能にしている。

また Crossref と DataCite が共同で進めている Event Data<sup>注10)</sup> というプロジェクトでは、Twitter や Wikipedia 等の様々な情報源における DOI 利用を調べ、そういった異なる情報源での識別子と DOI の関係を集積、提供することをやっている。

オーストラリアの Australian Research Data Commons (ARDC) が進める RAiD (Research Activity Identifier)<sup>注11)</sup> では、研究の個別の活動を多様な識別子のセットとして記述して、それに新たな識別子を与えるものである。一つの RAiD ID に組織や助成団体、研究協力者、利用しているツールやサービス、データの識別子が紐づけられる。

## 4. おわりに

学術ジャーナルの電子化およびそれに引き続くオープンアクセス化により、学術のオープン化、すなわちオープンサイエンスが進んできた<sup>注12)</sup>。これまで学術コンテンツは分野ごとに限られた人々だけがアクセスすればよいものであった。このためアクセス手段やそのコンテキスト (作成した著者や組織といった情報) はその人たちだけで共有すればよかった。しかし、オープンサイエンスにおいては、FAIR 原則<sup>6)</sup> に要約されるように、だれでもアクセスでき、だれでもそのコンテキストが理解できることが必要とされる。その実現のためには永続的な識別子によるアクセスと

様々な識別子を用いたコンテキスト情報の記述が必要となる。すなわち、オープンサイエンスにおいて永続的な識別子は必須の要素である。

DOI は運用開始より約 20 年経った。DOI は汎用の識別子で学術以外の分野に適用されてきたが、特に学術分野での DOI 利用において大きな成功を取めた。ジャーナル論文においては DOI の付与はデファクトスタンダードになり、さらに多様な学術コンテンツにも DOI が付与されるようになった。さらに DOI 運用の安定性から、DOI を用いて特定の対象に対する識別子を作る例もでてきている (例: 前述の Open Funder Ref)。すなわち、DOI は今後のオープンサイエンスを支える重要な基盤であり、かつ今後もより多様な展開が期待される。

最後に、DOI がこれまで運用できているのは、研究者や学術関係機関による貢献が大きいということを述べて終わりたい。研究者は積極的に論文内の引用文献やアクセス手段として DOI を使うことでその価値を高めているが、研究者の役割はそれだけではない。自らの学術コンテンツを公開するときに DOI の付与を選択することで、DOI 付与機関の活動を支援している。それは、結果としてそういった機関をメンバーとする RA 組織の安定的な運用に寄与している。永続的な識別子を基盤とするオープンサイエンスのエコシステムを維持発展させるためにも、今後とも研究者および学術関係機関の積極的関与を期待したい。

### 注・参考文献

- 注 1) Crossref 設立の経緯については 2) を参照のこと。
  - 注 2) <https://figshare.com/>
  - 注 3) ジャパンリンクセンター ストラテジー 2017-2022 [https://japanlinkcenter.org/top/about/about\\_strategy.html](https://japanlinkcenter.org/top/about/about_strategy.html) (参照 2021-02-12)
  - 注 4) DOI の ISO 規格である ISO 26324:2012 Information and documentation – Digital object identifier system においては kernel metadata が規定されているが、必須事項以外は活用されていない。
  - 注 5) <https://citation.crosscite.org/>
  - 注 6) <https://opencitations.net/index/coci>
  - 注 7) <https://www.crossref.org/services/funder-registry/>
  - 注 8) Open Funder Registry は DOI ネームを使っているので DOI 連携の一つとも言えるが、学術情報としては異なるものなのでここで紹介している。
  - 注 9) <https://commons.datacite.org/>
  - 注 10) <https://www.crossref.org/services/event-data/>
  - 注 11) <https://www.raid.org.au/>
  - 注 12) ここでは単純化して述べたが、オープンサイエンスの出自はもう少し複雑である。詳しくは 5) 等を参照。
- 1) 宮澤彰. 識別子序論. 情報の科学と技術. 2020, vol.70, no.5, p.262-265. [https://doi.org/10.18919/jkg.70.5\\_262](https://doi.org/10.18919/jkg.70.5_262)
  - 2) Crossref. “The Formation of CrossRef: A Short History”. 2009. <https://www.crossref.org/pdfs/CrossRef10Years.pdf> (日本語訳: 時実象一. CrossRef 誕生小史. 情報の科学と技術. 2010, vol.60, no.7, p.289-295. [https://doi.org/10.18919/jkg.60.7\\_289](https://doi.org/10.18919/jkg.60.7_289)). (参照 2019-02-12)
  - 3) 波羅仁ほか. DOI と JaLC の活動について. 情報の科学と技術. 2020, vol.70, no.8, p.428-431. [https://doi.org/10.18919/jkg.70.8\\_428](https://doi.org/10.18919/jkg.70.8_428)
  - 4) Peroni, S.; Shotton, D. OpenCitations, an infrastructure organization for open scholarship. Quantitative Science

Studies. 2020, vol.1, no.1, p.428-444.  
[https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023)

- 5) 武田英明. オープンサイエンスの成り立ちと学術コミュニケーションの未来. 情報処理. 2019, vol.60, no.5, p.392-398.
- 6) FORCE11. “THE FAIR DATA PRINCIPLES”. FORCE11.

2016. <https://www.force11.org/group/fairgroup/fairprinciples>, NBDC 研究チーム (訳). “FAIR 原則 (「THE FAIR DATA PRINCIPLES」 和訳)”. 2019.  
<https://doi.org/10.18908/a.2019112601>

**Series:** Identifiers used in information science and technology, 12: DOI as a scholarly persistent identifier: the story so far and future. Hideaki TAKEDA (National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430)

**Keywords:** DOI / persistent identifier / open science