

SPARC Japan セミナー2024

「オープンアクセス義務化の先にあるもの：来るべき世界に向けて」

知の循環のミッシングリンク： 知的資産はどのような利用事例を生み出すか？

北本 朝展

(国立情報学研究所)

講演要旨



知的資産へのアクセスを広く開くことで、知が人々の間で循環して利用され、さらなる知的資産の創造につながることを期待されている。しかし、知的資産の出版に関する情報の充実と比べると、知的資産の利用事例に関する情報は乏しく、ここが知の循環における「ミッシングリンク」となっている。そこで我々は、データセットの利用事例を収集・共有するプラットフォーム Mahalo Button の開発などを進めることで、ミッシングリンクを追跡可能とする試みを続けている。出版から利用へと知の循環が回る未来像を考察してみたい。



北本 朝展

国立情報学研究所コンテンツ科学研究系教授。データサイエンス共同利用基盤施設人文学オープンデータ共同利用センター長。

https://www.nii.ac.jp/faculty/digital_content/kitamoto_asanobu/

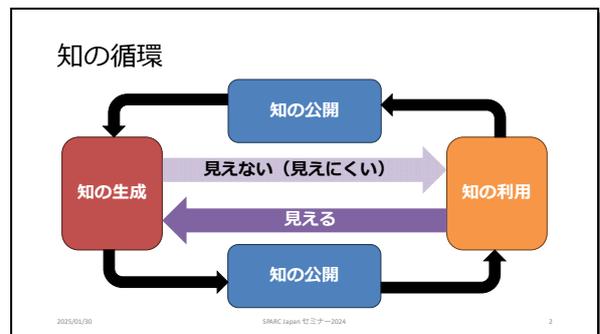
知の循環とミッシングリンク

図1は、知が生まれて公開され、それを利用し、また新しい知の生成につながるという循環を「知の循環」と名付けて図に表したものです。本日は、この「知の循環」についてお話しいたします。

今、オープンサイエンスで主に議論されているのは、生み出した知をどのように公開するか、オープン化するかどうかという「知の公開」の部分です。「知の循環」は、公開した知が別の人に利用され、それがまた公開されてつながっていきます。「知の生成」が「知の利用」をしたい人から見えることによって、知が使えるようになるというのがオープンサイエンスの一つのメリットです。ところが、知を生み出した側からは、それが

どう使われているのがよく見えません。あるいは、見えづらくなっています。ですから、知の循環とは言っても、実は知の利用・公開・生成はつながっていないのではないかと考えています。

「知の生成」から「知の公開」のところは、オーブ



(図1)

ンサイエンスの中心的な課題として、オープンアクセス、オープンデータなど多面的な活動が進んでいます。また、公開された知をどのように利用するかという、いわゆる再利用も、オープンサイエンスの一つの課題として議論されています。ところが、「知の利用」「知の公開」をしても、それが次の「知の生成」につながるどころが見えにくい。これを私は「ミッシングリンク」と呼んでいます。

生成側から利用側が見えれば、生成の価値が分かります。例えば、何本の論文に引用されたかを表すサイテーション（被引用数）は価値の可視化の一つです。さらに重要なのは、どのように使われたかという情報が集約されると、利用方法を汎用化できるということです。逆に利用方法に関する情報がバラバラになっていると、知がどのように使えるのかということがよく分かりません。

知の循環の課題

この循環の課題として、まず、知を作っても適正に評価されないということがあります。引用プラットフォームによる解決が議論されていますが、これは道半ばです。

また、知を公開しても、それがどう使われているのかよく分からないという悩みがあります。これは知の生成者だけでなく、実は知の公開者も同様の悩みを抱えています。リポジトリで公開しても、それがどのぐらい使われているか、誰が使っているかが分からない。利用事例が把握しづらいという悩みは、いろいろな人が持っています。

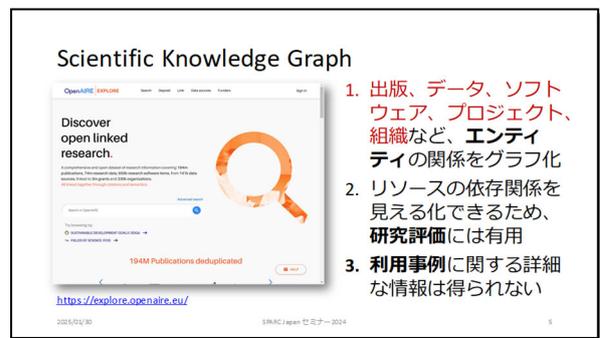
知を利用した人が、知の生成者や公開者にフィードバックをしたいと思っても難しいという課題もあります。例えば、改善の提案、あるいは単純に「ありがとう」と伝えることさえなかなかできないのが現状です。

そして、「知の循環」が切れたままだと、「知のFlywheel（弾み車）」を加速することができません。そのため、ポジティブフィードバックによる知の成長が起りにくいという問題があります。

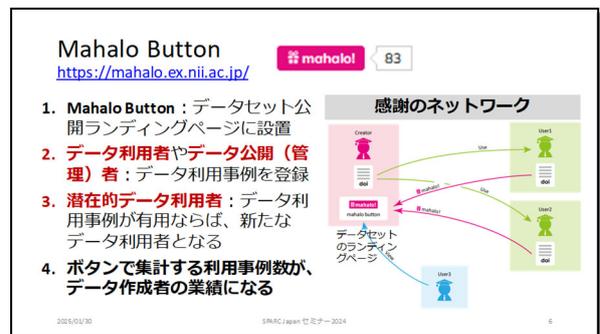
そこで、知をつなげるために活用されているのが、Scientific Knowledge Graph です（図 2）。これは、出版やデータ、ソフトウェア、プロジェクト、組織など、研究に関するエンティティの関係をグラフ構造としてリンクしようという考え方です。個々の研究リソースの依存関係が見える化できるため、研究評価などには使えます。ただし、単にリンクしているだけで、どう使われたかという利用事例に関する詳細な情報を得るのが難しいという課題があります。

DIAS と Mahalo Button

われわれの Mahalo Button は、このような問題を解決するためのサービスです（図 3）。Mahalo Button は、いわゆる「Like Button（いいね！ボタン）」と基本的な機能は同じで、ボタンを押すと情報を入力できるというものです。これをデータセット公開ランディングページに設置して、データ利用者やデータ公開者がデータ利用事例を登録していくことで、利用事例がたまるという仕組みです。これからデータセットを使おうとする人がそれを見ると、どのように使われているか



(図 2)



(図 3)

が分かり、利用事例から学ぶことによって新たなデータ利用者となります。

Mahalo Button は、既にデータ統合・解析システム (DIAS) のデータ公開ページに設置し、情報を入れています。図 4 の左下に「このデータセットを利用した論文」と表示されている「83」という数字は利用事例数で、これはデータ作成者の業績にもなります。これをクリックすると、図の右側にあるようにどのように使われたかが一覧として表示されますので、そのデータの使い方を知ることができます。

このサービスに最近、幾つかの改良を加えています。一つはデータ引用論文の検索支援です (図 5)。データセットがどう使われたかについては、デジタルオブジェクト識別子 (DOI) をキーとして検索すると、引用している論文にアクセスできるのが理想です。DataCite や OpenCitations の API を使うと、データセットが使われている論文が分かります。そこで「Give Card」ボタンを押すと、データ引用論文の登録が簡単にできます。

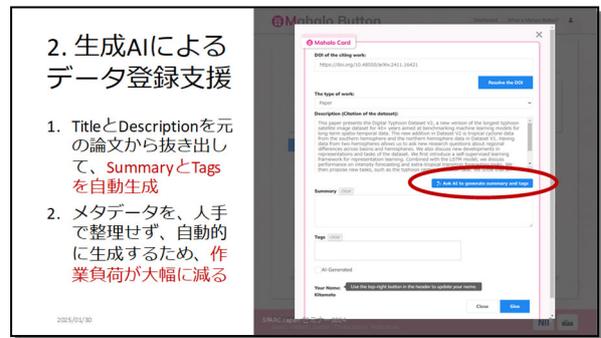
もう一つは、生成 AI によるデータ登録支援です

(図 6)。データセットがどのように使われているかということは、さまざまな形式で論文中に書かれています。そこで、「Title」と「Description」を入力することで、それを元に「Summary」や「Tags」を元の論文から抽出して自動生成します。このような生成 AI の活用も、情報の入力に関わる手間を減らす上で非常に重要です。

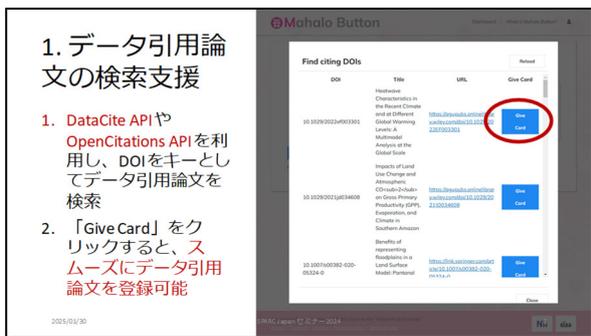
三つ目は、生成 AI によるデータ利用支援です (図 7)。例えば Mahalo Card に登録した利用事例に関して、「このデータセットの利用事例をまとめてください」と生成 AI のチャットに入れると、情報をまとめて「このデータセットはこんな形で使います」と結果を返してくれます。また、「気候変動への利用をもう少し詳しくまとめてください」とプロンプトを変えると、気候変動に関する利用事例が抽出されるなど、さまざまな情報を取得できます。このように、データセットの潜在的利用者が、どのように使えるかを把握してから、使うことができるような支援となっています。



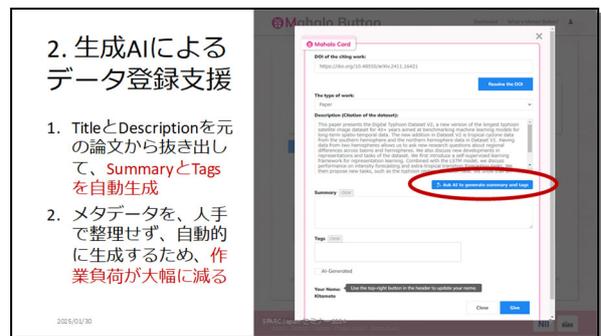
(図 4)



(図 6)



(図 5)



(図 7)

データ利用事例収集の課題

オープンサイエンス的には、データ利用者が全ての引用に DOI を明示し、それが自動的にデータベースに整理され、API 経由でオープンに利用事例が取得可能となることが理想です。ただ現実には、DOI の利用は限定的で、データへの言及方法がバラバラですので、API だけでは利用事例を網羅できません。そのため、さまざまな手段を使って事例を収集する必要がありますが、これを行うのは大変です。

理想は、ユーザー全員が「使いましたよ」と Mahalo Button にボランティアに登録することで、データベースが充実していくという方向です。しかし実際のところ、ボランティアな登録はなかなか難しいため、われわれ DIAS では現在、リポジトリ側の付加価値となるよう、リポジトリ側で情報を登録しています。利用事例の収集は、リポジトリの評価と運営においても、付加価値として広く使えるのではないかと考えます。

Mahalo Button ご利用のすすめ

Mahalo Button は DIAS プロジェクトで開発・利用を進めていますが、DIAS 専用のシステムではなく、誰でも自由に使うことができます。

データ管理者としては、利用事例を共有することでデータに付加価値を与え、利用を増やすサービスとして利用できます。データ提供者・公開者としては、データの利用事例を把握し、貢献を可視化するサービスとして利用できます。さらにデータ利用者は、生成 AI チャットなども使いながら、データの利用方法を把握し、自らの研究に使うことができると考えています (図 8)。

ご利用のすすめ

<https://dias.ex.nii.ac.jp/mahalo/>

1. Mahalo ButtonはDIAS専用のシステムではなく、GoogleのFirebase認証により、誰でも使えます
2. データ管理者：利用事例の共有によりデータに付加価値を与え、利用を増やすサービスとして利用
3. データ提供者：データの利用事例を把握し、貢献を可視化するサービスとして利用
4. データ利用者：生成AIチャットの支援によりデータの利用方法を把握し、自らの研究に利用

2025/01/29 SPARC Japan セミナー2024 12

(図 8)