

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」

学術情報流通の次の10年の見取り図

武田 英明

(国立情報学研究所)

講演要旨



オープンサイエンスの進展によって学術情報流通の世界は大いに変貌しつつある。これまでは論文のみによるシンプルな学術情報流通であったが、多様な情報源で公開される、論文・研究データ・ソフトウェアといった多様な研究成果を扱うようになっている。また利用の用途も研究だけでなく研究評価までに広がっている。このような学術情報流通の見取り図を示すとともに、その中で中心を占める PID の役割について述べる。



武田 英明

国立情報学研究所情報学プリンシプル研究系教授／主幹兼知識コンテンツ科学研究センター長。

1991年3月東京大学大学院工学系研究科博士課程修了。工学博士。ノルウェー工科大学、奈良先端科学技術大学院大学を経て、2003年5月より国立情報学研究所教授（現在に至る）。2024年4月より同研究所知識コンテンツ科学研究センター長。知識共有、Web 情報学、設計学などの研究に従事。人工知能学会、電子情報通信学会、情報処理学会、精密工学会、AAAI各会員。

http://www.nii.ac.jp/faculty/informatics/takeda_hideaki/

私は長らく SPARC Japan に関わってきましたが、現在は少し立場が変わり、2024年4月より知識コンテンツ科学研究センター長を務めております。このセンターは CiNii の開発を担当するセクションです。本日は実際にデータをハンドリングする立場から、学術情報流通のこの先10年がどうなるのかという話をしたいと思います。

学術コミュニケーションの変化

今、学術コミュニケーション、学術流通は大きく変化しており、いわゆるビッグデータ化が進んでいます。3V で見ても、規模（volume）は拡大し、毎年指数級に増えています。また、多様性（variety）が広がり、書籍、雑誌論文、国際会議論文だけでなく、プレプリント、研究データ、ソフトウェアなども研究成果とし

て把握しなければならなくなっています。また、スピード（velocity）も加速し、競争が激しい分野は、年のスケールから月のスケール、さらには日のスケールへと変化しています。プレプリントで有名な arXiv には機械学習系の論文が数多く公開されますが、例えばある日に投稿された論文が、次の日の arXiv の論文に引用されるような時代になってきています。

また、学術コミュニケーションに対して、より多くの要請がなされるようになっていきます（図1）。今までは研究者の間できちんとやっていればよかったのですが、根拠の提示や再現性の追求などがより強く求められるようになりました。査読についても、制度の検証や新たな制度の模索、オーサiershipについても著者の役割の明示化などの再定義・再定置が求められています。さらに研究資金の明確化や、研究組織のコミ

ットメントも求められるようになっていきます。

このような中で、学術コミュニケーションは一体どこに行くのでしょうか。一つは、オープン化です。これは間違いない道で、特にサイテーションを含む研究成果情報（メタデータ）は基本的にオープン化されていきます。今までサイテーションはかなりクローズで、Crossref も全部は出していませんでしたが、OpenCitations でほとんど出るようになりました。また、研究成果自体もオープン化しています。これがオープンサイエンスのコンテキストです。

次に、永続識別子（PID）化です。先ほど述べたように、データ量が増大し、大量の研究成果がありますので、きちんと ID を付けないと管理できない時代になりました。研究成果・論文だけでなく、研究者、組織、研究助成金課題などの ID 化も必要です。

そして、データ利用の高度化です。単に検索できればよいわけではなく、分析や総合という使い道となっていくと思います。おおよそその全体の方向性をこのように考えています。

それぞれのステークホルダーの役割

このような状況の中で、さまざまなステークホルダーからいろいろな要請が来ています。それぞれの役割について簡単にお話しします。

研究者の役割は、研究助成金課題を申請し、研究活動を実施し、研究成果を公開することですが、その役割はさらに増えています（図2）。



学術コミュニケーションの変化：より多くの要請

- 根拠の提示、再現性の追求
 - 根拠データの公開
 - 関連データやソフトウェアの保存
- 査読の再定義、再定置
 - 査読制度の検証
 - 新しい査読制度の模索
- オーサシップの再定義、再定置
 - 著者の役割の明示化
 - 著者所属
- 研究資金の明確化
 - 研究の公平性
- 研究組織の役割の再定義、再定置
 - 研究への責任の明確化

（図1）

また、研究機関・大学の役割も増えています。単に機関リポジトリを運営するだけでなく、所属する研究構成員の研究成果の把握も機関の役割となっています。大規模大学では対応もできますが、中小規模の大学の場合、それら全てを少人数で対応しなければならないため、大変な事態になります（図3）。

助成機関は、今までは資金を提供するだけでよかったのですが、研究助成金課題の成果をきちんと公開しなければならなくなっています。公的資金の場合は特にそうです。研究成果情報の公開なども助成機関の役割となっています（図4）。

さらに政府も、研究助成制度全体を監督する立場がより強化されています。府省共通研究開発管理システム（e-Rad）なども強化し、システムのサポートをしようとしています。政府も責任を負うことになります。そして、これらのデータを収集してevidence-based



研究者

- 役割
 - 研究助成金課題の申請
 - （研究活動の実施）
 - 研究成果の公開
 - 研究助成金課題の報告書の提出
 - 関連研究成果の提出
 - 研究成果の大学への登録
- 現在の状況
 - 成果の公開と大学へ登録、助成機関への報告が連動小
 - 組織的サポートの不足
- 将来の展開
 - 研究データ基盤システムによるサポート

（図2）



研究機関(大学)の役割

- 役割
 - 機関リポジトリ(IR)の運営
 - 大学発の研究成果の発信
 - 個別のデータリポジトリの運営
 - 研究センター等がデータを公開
 - 研究情報システム(CRIS)の運営
 - 所属研究者やプロジェクトの情報の発信
 - 研究情報管理システム(RIMS)の運営
 - 機関の意思決定のための自機関の研究の分析や報告
- 現在の状況
 - IRはほぼ整備、シンプルなCRIS相当のサービスはあるが、無秩序。RIMSはほとんど導入されていない
- 今後の展開
 - IR<->CRIS/RIMSの相互運用の拡大
 - 大規模大学でのパッケージRIMSの導入。RIMSがCRISを兼ねる
 - 中小大学でのCRISの"標準化"

（図3）

policymaking (EBPM) に使うこともしています (図5)。

それから、CiNiiなどの学術データベースです。これらの関連する学術情報を全て引き受けて、網羅的体系的に収集し、ユーザーに提供する役割があります (図6)。

これが、現在進行中の学術コミュニケーションの変化です。

支援システムの現状とシステム設計の原則

現状、図7のようにさまざまな支援システムがありますが、一体どのようなシステム設計を考えればよいのでしょうか。

設計の原則の一つは、公開系と事業系のシステムをきちんと区別することです。オープン化は大変重要ですが、一方で、科学は全てがオープンというわけではありません。共同研究もありますし、大学にはビジネ

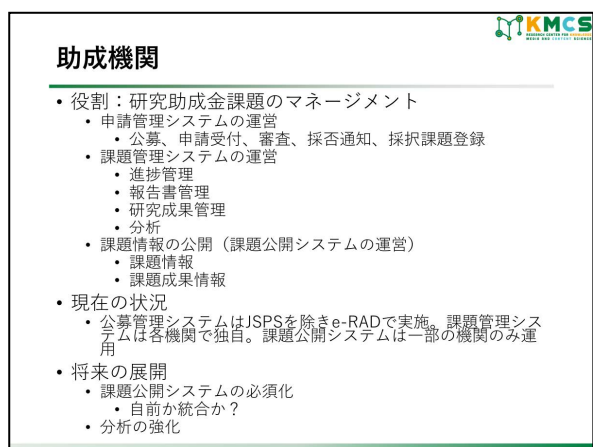
スもあります。それらは事業系のものとして区別します。

二つ目が、主体の役割の明確化です。誰が管理するのか主体をはっきりさせるということです。

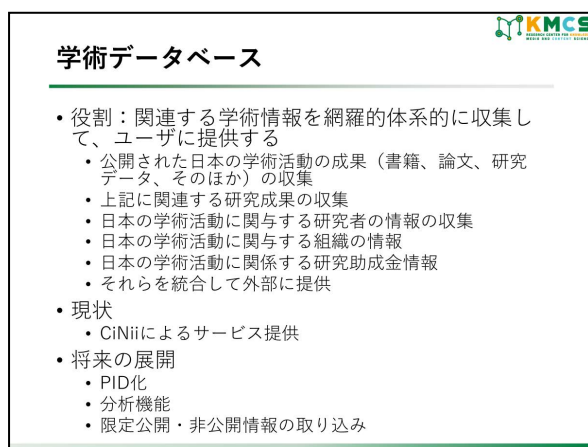
三つ目が、多重系です。これだけ多様なステークホルダーがいますので、どんな方法でも必ず漏れがあります。いかにシステムを多重に作って、完全でないデータを相互に補うかということです。

日本の学術コミュニケーションの見取り図

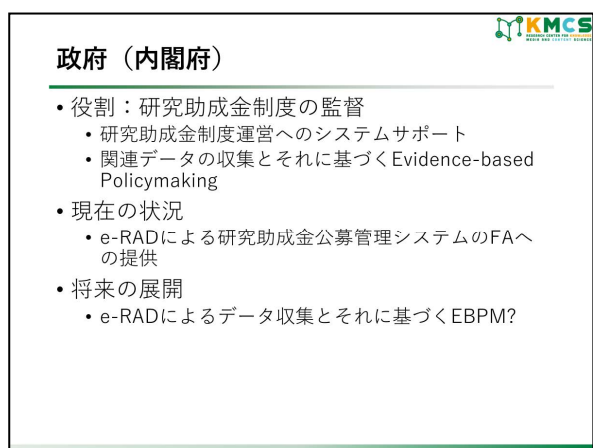
図8は、今後10年の研究成果情報の流れを示したものです。ステークホルダーとして、一番左に研究者・研究機関があります。下部は、学術コンテンツを公開している学会や出版社、サブジェクト・リポジトリなどです。上部は助成機関や政府、右側はリユーザーです。真ん中の右にある学術データベースは、まさ



(図4)



(図6)



(図5)



(図7)

にデータを収集して公開するサイトです。

研究助成登録課題情報の流れについては、色付けで示しています（図9）。緑色はコンテンツの流れです。つまり、論文などを書いたら、機関リポジトリに公開するか、学会サイトに公開するか、出版社サイトに公開することになります。青色はメタデータの流れです。本文とは別にメタデータが収集される点も、学術情報流通の重要なポイントです。メタデータの流れには、公開したサイトから出る、所属の大学から出る、あるいは助成機関の成果報告として出るという、大きく3通りのルートがあります。学術データベースはこれらのメタデータを全て集めなければならないこととなります。

非常に複雑な状況ですが、真ん中にPIDがあれば何とか成り立つだろうというのが、現在の見取り図です。PIDがIDを共有することによって、いろいろなステ

ークホルダーから出てくるデータをうまく特定して流すことができるのです。PIDの一つであるデジタルオブジェクト識別子（DOI）は、主に研究成果に付けます。それ以外に、研究者に付ける Open Researcher and Contributor ID（ORCID）、組織に付ける Research Organization Registry ID（ROR ID）、助成課題に付ける IDなどもPIDとしてシェアするものです。PIDが中心になれば、この複雑な仕組みも流れることとなります。

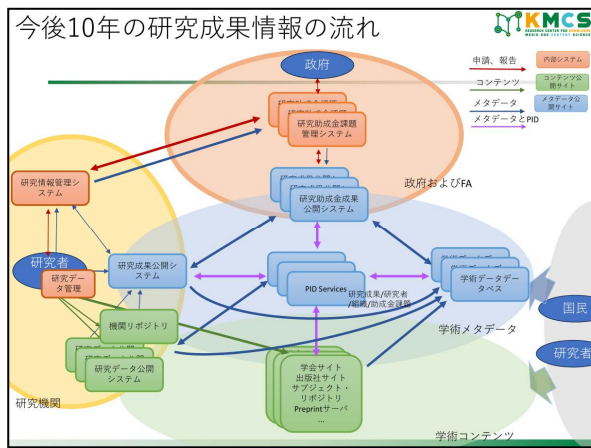
例えばメタデータに着目すると、研究助成金課題情報のIDをたどれば課題の成果報告として集約され、このような流れで最後に右側（学術データベース）に集まることとなります。

現在、日本で実際に動いているシステムを図の上にマッピングすると、図10のようになります。一番右側にある学術データベースのCiNiiにたくさんの線が向かっていることから、ここがいかに大変な状況か分かっていただけたと思います。これが現在進行中の日本の学術コミュニケーションの見取り図です。

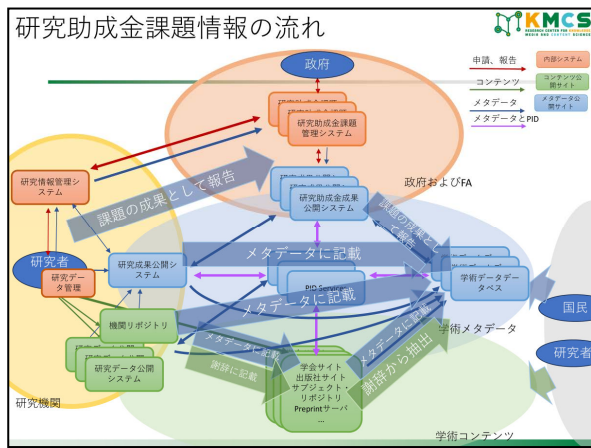
まとめ

現在、学術情報流通は、その量、多様性、スピードが増大しています。また、環境が複雑化しており、研究者間の情報のやり取りだけではなく、助成機関側、市民側、データを使う側という外部要因があり、さらに公正性という問題もあります。そのためさまざまな要請が増えています。

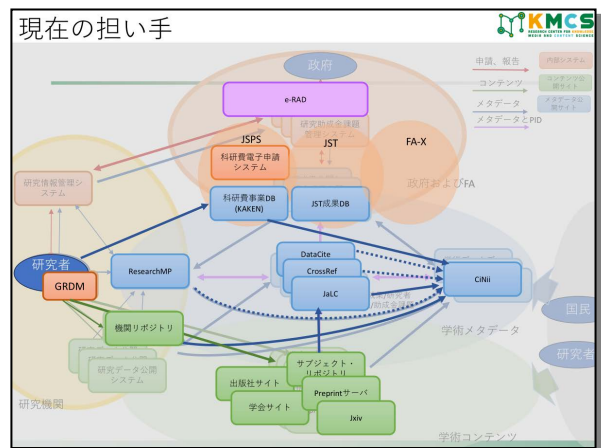
先ほど示した次の10年の見取り図は、とにかく多



（図8）



（図9）



（図10）

様なステークホルダーが絡み合う複雑なネットワークとなります。ここでPIDを核とすることで、コンテンツ、人、組織、課題が、個々のPIDがあるからこそうまくいくということになります。

一方で、他の方からもご指摘があったように、学術情報流通が増大すると人間が扱えなくなることが予測できますので、自動化や AI 化が加速するといえます。世界では既にPIDの発行競争が始まり、PIDベースのサービスが増えています。日本においては、誰がどのような役割でPIDを作成・利用するのかというPIDポリシーをまず制定しなければなりません。そして、PID発行の責任体制を構築する必要があります。これらが近々の課題だと思います。