

SPARC Japan セミナー2021

「研究データポリシーが目指すものとは」

研究 DX を巡る政策動向から見る 研究データポリシーの役割

三宅 隆悟

(文部科学省 学術基盤整備室)

講演要旨



第6期科学技術・イノベーション基本計画では、オープンサイエンスとデータ駆動型研究の推進が柱の一つとして位置づけられた。この方向性の下に、具体的な取り組みとして「公的資金における研究データの管理・活用にに関する基本的な考え方」がまとめられ、研究開発を行う機関の責務の一つとしてデータポリシーの作成が位置づけられている。本講演では、研究デジタル・トランスフォーメーション（研究DX）の推進に向けた政府の政策の動向を解説しつつ、その中での研究データポリシーの役割・位置づけや現状について紹介する。



三宅 隆悟

東京大学大学院情報理工学系研究科修士課程修了（情報理工学修士）。2003年文部科学省入省。原子力、海洋や競争的資金、研究開発基盤等の政策立案に従事。内閣府、内閣官房、復興庁への出向のほか、理化学研究所計算科学研究推進室ではスーパーコンピュータ富岳の設置に向けた企画調整業務を担当。2020年4月より現職。

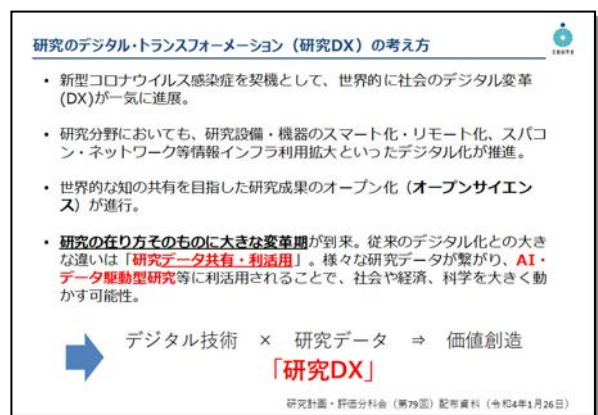
研究のデジタル・トランスフォーメーション （研究DX）の考え方

私からは、研究データポリシーの作成に当たって一番関連の深い研究DXの施策動向を、海外の動向等も踏まえながら紹介させていただきます。研究DXにもさまざまな議論があり、いろいろな定義がありますが、文部科学省としては以下のように定義して進めています（図1）。

新型コロナウイルス感染症（COVID-19）を契機として、社会におけるデジタル変革やDX化が進みました。研究分野においても、例えば研究機器のスマート化・リモート化、スーパーコンピュータ・ネットワーク等の情報インフラのデジタル化が進んでいます。一方、世界を見ると、やはり知の共有を目指した研究成果のオープン化（オープンサイエンス）が目指されて

いる状況があります。このような状況を踏まえると、研究の在り方そのものについても大きな変革が起きているのではないかと考えているところです。

オープンサイエンスやデータ駆動型研究を推進していくことによって、付加価値の高い研究成果を創出し



（図1）

てわが国のプレゼンス向上を目指していくためには、研究データの管理・共有・利活用が重要です。デジタル技術と研究データを合わせることによる価値創造を目指して進めていく必要があると考えています。

このような状況を踏まえて、研究データの共有・利活用によってデータ駆動型研究による研究プロセスの変革およびイノベティブな成果の創出を期待しているところです。

背景として、研究データや論文の提供が論文掲載の要件になったり、世界的な出版社やIT企業がビジネスの対象として興味・関心を高めていることもあり、研究データに関する取り組みは非常に活発化している状況です。

これを踏まえて、そもそも研究データというものが戦略的資源であって、その共有が論文執筆と同等以上の価値を持つという基本的な理念を共有していく必要があります。そのためには、国全体として研究データの共有・利活用ができる基盤・体制を整備し、その中で研究データの共有を促進してデータ駆動型研究を推進するという流れをつくっていく必要があると考えています。これらを早期に起こして動かしていくことで、国際競争力・協調力の向上を図ることを目指すわけです。

情報科学技術が支える学術情報基盤（研究開発デジタル基盤）

もう少し広い観点で言うと、情報科学技術全体で学

術情報基盤をしっかりと支えていく必要があると考えています（図2）。研究データ自体を管理する研究データ基盤の構築・高度化という観点もありますし、「富岳」を代表とする国内のスーパーコンピュータをしっかりとつないで計算のリソースを用意する。また、つなぐためにはネットワークが必要なので、学術情報ネットワーク SINET を整備して、これらを一体的に活用して学術基盤を整備し、例えばマテリアル、ライフサイエンス、環境・防災、その他の領域も含めて、研究データの管理・利活用やDXを進めていくということに取り組んでいます。図3は、これに関連する研究データプラットフォームの一例です。さまざまな分野において、さまざまなデータプラットフォームが存在しています。こういうものをつなげて研究を加速していくことを目指しています。

オープンサイエンスに係る世界と日本の動向

論文のオープンアクセスと研究データのオープン化を含む概念であるオープンサイエンスは、2013年のG8による「オープンデータ憲章」の合意を発端に世界的な潮流となっています。また、それ以外のさまざまな国際機関においても議論が行われ、FAIR原則に基づくデータの共有や、人材育成・供給などの国際的な対応が求められているところです。

ただ、もちろんオープンといっても全てを公開すればいいのかというと、当初はそういう議論も若干あったと思いますが、そういうわけではありません。例え



(図2)



(図3)

ば個人情報や秘密情報、特許のような権利的な話や、
わが国にとって重要な情報を含むものなどに関する公
開・非公開については、しっかりと戦略的に行う必要
があります。

改めて海外の動向を見ると、オープンアクセスの急
速な普及に伴って、例えば欧州では Plan S や
Horizon2020 のような取り組みでオープンアクセスの
義務化の流れが起こっていたり、さまざまな分野で論
文のエビデンスとなる研究データについて公開が投稿
条件になったりしています。それに合わせて海外大手
出版社が論文出版にとどまらず、研究データを含むさ
まざまな情報を支えるプラットフォームに拡大してい
る状況です。

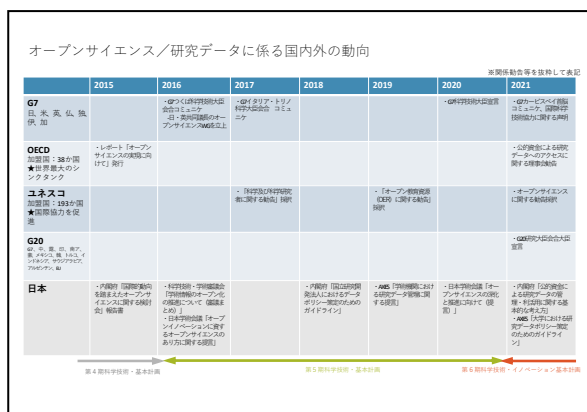
具体的な議論を紹介します。2021 年の G7 における
オープンサイエンスの会議の後、G7 研究協約
(Research Compact) の中で「データ及びツールを可
能な限りオープンかつ迅速に共有する際の障壁を取り
除くとともに、オープンサイエンスを促進し、オープ
ンで安全かつ透明性のある形で市民への科学の普及を
拡大させ、技術関連のリスクの最小化に努める」とい
うことが掲げられました。こういうものが首脳レベル
での議論の中で出てきたということです。

それを時系列で整理したものが図 4 です。もちろん
これに至るまでには、2016 年の G7 茨城・つくば科学
技術大臣会合をはじめ、さまざまな議論が行われてい
ます。当然、日本においても内閣府や、文部科学省で
は科学技術・学術審議会、また日本学術会議等々で議

論が行われて、今の政策の状況につながっているとい
うことです。

科学技術・イノベーション基本計画と、基本 的な考え方

ここからは政策の中でどのような位置付けがなされ
ているかを紹介させていただきます。図 5 は、第 6 期
科学技術・イノベーション基本計画の概要です。この
中にさまざまな要素が含まれていますが、「新たな研
究システムの構築（オープンサイエンスとデータ駆動
型研究等の推進）」という章が設けられました。基本
計画上も、この取り組みに注力していくということが
表されていると考えます。ここでは、研究データの管
理・利活用、スマートラボ・AI 等を活用した研究の
加速や機器の共用、研究 DX が開拓する新たな研究コ
ミュニティや環境の醸成をしっかりと進めていこうと
いうことが明確にされています。取り組みとしてはさ
まざまな観点から行うこととなっていますが、特に本
日の議論に関係するのが、「公的資金により得られた
研究データの機関における管理・利活用を図るため、
大学、大学共同利用機関法人、国立研究開発法人等の
研究開発を行う機関は、データポリシーの策定を行う
とともに、機関リポジトリへの研究データの収載を進
める。あわせて、研究データ基盤システム上で検索可
能とするため、研究データへのメタデータの付与を進
める」という部分です。こういった取り組みを基本計
画上では求めています。



(図 4)

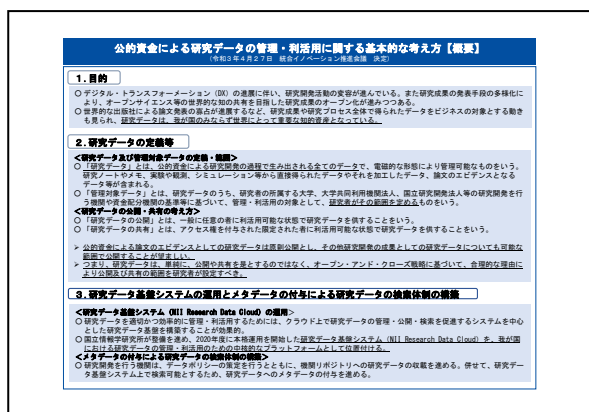


(図 5)

それを具体的に進めるために、2021 年 4 月に、閣僚級会議である統合イノベーション戦略推進会議で「公的資金による研究データの管理・利活用に関する基本的な考え方」が決定されています。この中では、研究データ基盤システムを整備し、これを中核的なプラットフォームとして、研究データの管理・利活用に向けた取り組みを推進していくということが書かれています。さまざまなデータプラットフォームとの連携を推進し、また大学・研究機関等においては研究データにメタデータを付与して収載していく、資金配分機関においても公募型の研究資金の中にその考え方を入れていくということが定められています。

「基本的な考え方」において目指す主要な数値目標としては、機関リポジトリを有する全ての大学等において、2025 年度までにデータポリシーの策定率が 100%になること、また、公募型の研究資金の新規公募分について、2023 年度までにデータマネジメントプラン（DMP）およびそのメタデータの付与を行う仕組みの導入率が 100%になることが掲げられて、今、進められている状況です。

もう少し詳細をご紹介します（図 6）。「基本的な考え方」は、最初に目的をしっかりと書いており、そもそもなぜこういうことをするのかという理念を表現しています。「1.目的」の部分では、研究データはわが国のみならず世界にとって重要な知的資産となっているという考え方を基に、さまざまな取り組みを進めていきたいと思います。



(図 6)

「2.研究データの定義等」では、研究データや管理対象データの定義をしています。ここで大事なのは、研究データそのものという、研究開発の過程で生み出された全てのデータを指しますが、当然それを全部保管したり管理したりするには手間がかかり、無駄も多いです。そのため、管理対象データを決めて、その範囲でしっかりと管理していこうというのがここでの考え方です。

その中で、研究データの公開・共有の考え方として、「公開」とは、一般の方にも見ることができるよう共有化の形で研究データを供すること、「共有」とは、アクセス権を付与された限定された人に研究データを供することとしています。ここでは公的資金の話による研究データのことも整理していますので、公的資金による論文のエビデンスとしての研究データは原則公開として、その他についても可能な範囲で公開するのが望ましいとしています。やはり研究データは、単純に公開や共有をすることが目的ではなく、合理的な理由により公開・共有範囲を研究者が設定して、しっかりと管理しようということがここで整理されています。当然、これをいったからといってできるわけではありませんので、さまざまな取り組みを合わせて進めていくことを考えています。

「3.研究データ基盤システムの運用とメタデータの付与による研究データの検索体制の構築」には、NII Research Data Cloud を中核的なプラットフォームとして位置付け、しっかりと運用していく、開発していくことを書いています。

また、各ステークホルダーにおける責務、すべきことを書いているのが図 7 です。「6.研究者の責務」では、研究者が管理対象データの決定、メタデータの付与、DMP の作成を行い、保存し、オープン・アンド・クローズ戦略に基づいて公開・共有することを求めています。

それを行うための機関としての責務を「4.研究開発を行う機関の責務」にまとめています。データポリシーの策定や、研究データへのメタデータの付与の推進、

それを支援する体制の整備・評価などです。

また、公募型の研究資金における資金配分機関（FA）の責務については 5 に記載しており、DMP やメタデータの項目の決定、これらの取り組みの評価体系の導入、e-Rad への登録を求めています。

オープンサイエンスの話を中心に説明しましたが、7 にあるように、これらの取り組みは研究公正の観点からも非常に重要です。近年、学术论文の査読に当たって、研究公正の観点から研究データの提供やリポジトリでの保存等が求められる場合があります。当然、研究公正の確保は機関や研究者にとって非常に大事なことです。これらをしっかりと説明できるようにするためにも、研究データはデータポリシーに従って適切に保存・管理される必要があります。

研究データ基盤システム（NII Research Data Cloud）

これらに関する施策をご紹介します（図 8）。研究データ基盤システム（NII Research Data Cloud）は、データ管理基盤である GakuNin RDM、検索基盤である CiNii Research、公開基盤である JAIRO Cloud の三つの基盤から構成されており、オープン・アンド・クローズ戦略に対応した研究データ基盤として開発されています。当然ながら、今後管理する上での負担軽減も進めていく必要があるため、例えばデータマネジメントプランに基づくデータ管理支援や大学等研究機関の情報管理といった負担軽減や、研究活動の支援がしや

すくなるような改良・高度化を文部科学省としてもしっかりと支援していきたいと考えています。

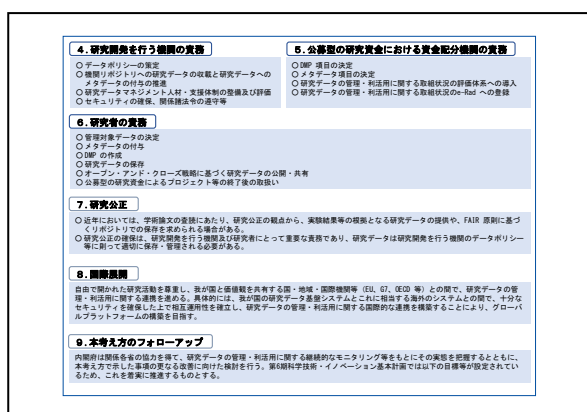
研究データポリシー策定のためのガイドライン整備状況について

二つのガイドラインをご紹介します。元々のデータポリシーの策定については、2018 年の「統合イノベーション戦略」において求められた関係で研究開発法人が先行して動いています。そして 2018 年に「国立研究開発法人におけるデータポリシー策定のためのガイドライン」が定められて、取り組むべき事項や定めるべき項目が列挙されました。

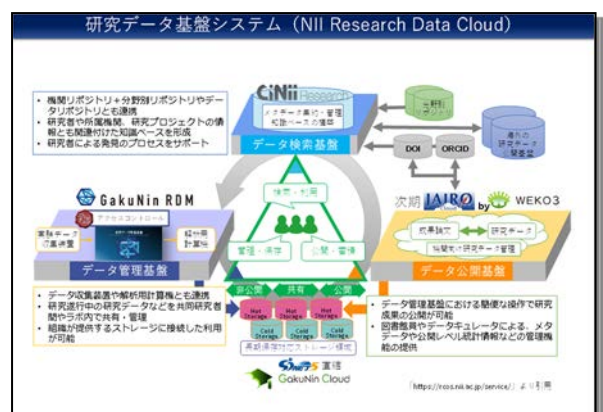
大学関係については、領域が広く、また自主性の問題もあり、国におけるガイドライン策定はなかなか進んでいませんでしたが、2021 年に大学 ICT 推進協議会（AXIES）で「大学における研究データポリシー策定のためのガイドライン」を策定していただきました。ここでは、研究データポリシー策定に当たっての留意点等々が整理されています。

データポリシーとデータマネジメントプラン（DMP）

研究開発法人につきましては、全法人についてデータポリシーの策定を目指すということが以前から行われており、文部科学省所管の 7 法人は全て策定済みです。大学の方で先行して策定された事例も把握しています。これらの整備状況も参考にしながら、ぜひ各機



（図 7）



（図 8）

関においてマネジメントプランの検討を進めていただければと思っています。

また、資金配分機関、競争的資金の関係でも順次検討を進めています。図9は文部科学省における研究データマネジメントプランに関する検討状況です。文部科学省が直接公募を行う各競争的資金制度においても、研究データマネジメントプランに関する記載の追記を検討しており、既に追記された公募も先行して行われています。こちらにつきましては、当室と競争的資金の関係を行っている文部科学省の関係部署とも具体的な書きぶりについて相談して、2023 年以降は、基本的に新規の公募についてはこのような記載が入る形で準備を進めているところです。

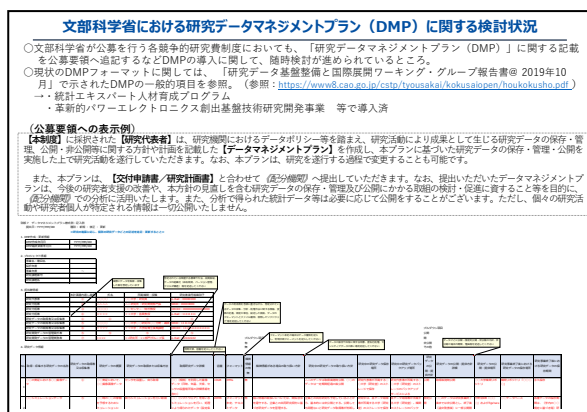
また、資金配分機関でも、先行して動いている例があります。科学技術振興機構（JST）であれば、さきがけ、CREST 等で 2017 年度より実施方針について記載があります。日本医療研究開発機構（AMED）、新エネルギー・産業技術総合開発機構（NEDO）においても 2018 年よりこの記載がされています。日本学術振興会（JSPS）の科学研究費補助金（科研費）の関係では、2021 年度の学術変革領域公募において明示しています。その他の内容についても現在検討中と聞いています。

研究データ公開のポジティブな影響

最後に、研究データ公開に関してのポジティブな影響を一つご紹介します。図 10 は、科学技術・学術政

策研究所（NISTEP）が実施した日本の研究者によるデータ・論文の公開状況への認識を明らかにするための調査（研究データ公開と論文のオープンアクセスに関する実態調査 2020）のうち、研究データを公開することによって得られるポジティブな影響に関する調査結果を抜粋したものです。今回モデレーターを務めていただいている林先生、池内先生を中心に行った調査と聞いていますが、簡単に紹介させていただきます。左の円グラフを見ると、研究データを公開した経験がある方の中で、公開によって良い影響があったと回答した方が 23%いたということです。元々こちらの質問は自由記述ですので、研究データのメリットを確かに感じている研究者の方がいるということだと思えます。右側の表は、研究データの公開の良い影響として回答された内容の詳細です。例えば、共同研究や交流の契機になった、引用の増加や認知度の向上につながったという回答があり、研究データの公開は研究活動にとって確かに良い影響があると感じている先生方もいるということだと思っています。

ここまで、いろいろな内容をご紹介してきましたが、研究データの管理・利活用に向けてさまざまな施策を打っています。文部科学省としても、その管理・利活用がしっかりとできるような形で基本整備を進めたいと思います。各機関、各研究においてこのようなことに取り組んでいただき、さまざまな形で研究データの管理・利活用を進めていただければと思っておりますので、ご協力のほどよろしくお願いいたします。



(図 9)



(図 10)

◆ 矢吹 寄せられた質問を読み上げさせていただきます。「諸外国と比して日本はデータ利活用の面でも既に出遅れているのではないかと漠然と懸念します。日本の学术界が魅力的なものになるためには、特にどのような政策あるいは研究データポリシーが重要とお考えでしょうか。お立場からの回答ではなく、個人的見解でも構いません」。いかがでしょうか。

◆ 三宅 こちらの研究 DX も含めて、さまざまな取り組みを先行して進められている例はあるかと思います。諸外国においても、やはり具体的な取り組みをしていくことは非常にハードルが高く、研究領域によって、進んでいるところもあれば、なかなか進んでいないところもあるという状況を聞いています。全てを一括して動かしていくのは、このような研究の在り方の変容のようなことを含む話ではなかなか難しいと考えておりますので、まず、できるところから取り組んでいく。文部科学省としても基盤整備をしっかり行い、使い勝手のいい形を準備していくということもありますし、研究の中ではデータ駆動型研究に着目していただく。全員が全員やらなくてはいけないということはありませんが、着目した先生方がそれぞれその研究を進めていき、必要な整備を行っていく。もちろん時間が限られているので、ゆっくり取り組めばいいというものではありませんが、その繰り返しの中で一步一步、着実に実績を積み重ねて広めていく取り組みが非常に重要ではないかと考えています。