第 3 回 SPARC Japan セミナー2020

「初めての研究データ」

データリポジトリ J-STAGE Data でのデータ公開

加藤 斉史

（国立研究開発法人科学技術振興機構）

講演要旨

CC BY

国立研究開発法人科学技術振興機構（JST）は J-STAGE 利用者を対象とするデータリポジトリ J-STAGE Data のパイロット運用を令和 2 年 3 月より開始し、令和 3 年 3 月に本格運用に移行する予定である。J-STAGE Dada サービス概要およびパイロット運用状況について紹介する。

加藤 斉史

教育サービス会社でeラーニング教材の開発業務に従事した後、2000年に科学技術振興事業団に入職。現在、科学技術振興機構情報基盤事業部にて、J-STAGEおよびJ-STAGE Dataの運営業務に従事。

はじめに

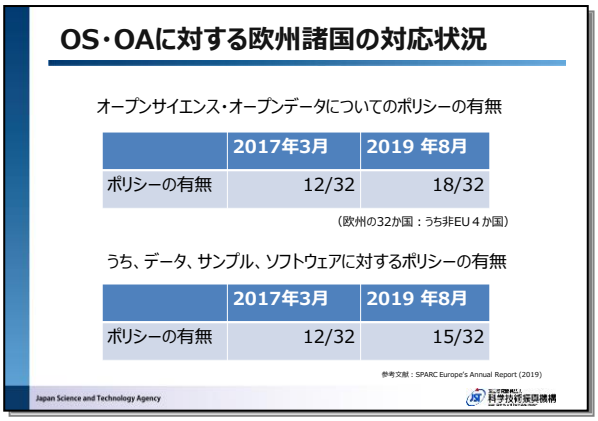
本日の話の内容は三つに分かれています。まず、J-STAGE Data 導入の背景として、データ共有が広まってきている状況についてお話しします。続いて、J-STAGE Data の概要として、データの登載から公開までの流れ等について紹介します。最後に、現在までパイロット運用を行っていますので、その取り組みと結果について紹介します。

J-STAGE Data 導入の背景

研究データ公開が必要とされる背景には、オープンサイエンスやビッグデータを活用したデータ駆動型研究の潮流があると考えています。データの共有、利活用による新たな価値を創造する取り組みが期待されています。

その一例として、欧州におけるオープンサイエンスの対応状況について 2019 年に調査を行いました（図 1）。32 カ国のポリシーを確認したところ、オープンサイエンスに関するポリシーがある国が 2017 年には 12 カ国だったのが、2019 年には 18 カ国に増えています。このうち、データに関するポリシーがある国は、2017 年の 12 カ国に対し、2019 年には 15 カ国に増えています。

また、研究の不正防止や透明性担保のために、多くの大学や研究機関、研究資金助成機関等が、データの管理・公開に係る方針を策定しています。図 2 は、アメリカやヨーロッパ、アジアの代表的な助成機関の対応状況についてまとめた表（2019 年時点）です。10



OS・OAに対する欧州諸国の対応状況

オープンサイエンス・オープンデータについてのポリシーの有無

	2017年3月	2019 年8月
ポリシーの有無	12/32	18/32

（欧州の32か国：うち非EU 4 か国）

うち、データ、サンプル、ソフトウェアに対するポリシーの有無

	2017年3月	2019 年8月
ポリシーの有無	12/32	15/32

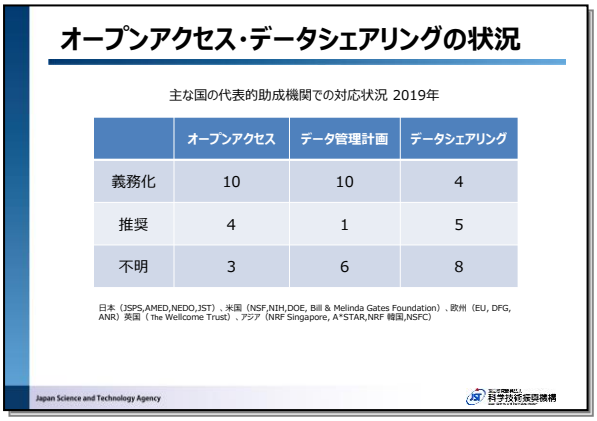
参考文献：SPARC Europe's Annual Report (2019)

Japan Science and Technology Agency

（図 1）

機関がオープンアクセスやデータ管理計画について義務化しており、進んできていることが見て取れます。データシェアリングについては、義務化 4 機関、推奨 5 機関と、少し強度がばらついています。これは、データの特性が分野によって異なることなどから、ポリシーの強度がばらついたのではないかと考えています。このように、データ共有についてもポリシーを定める機関が増えてきています。

次に、データの公開や共有に関するポリシーの整備がジャーナルでも進展していることについて紹介します。図 3 は、学術雑誌 220 誌に対してデータ共有ポリシーの強度の調査を行ったものです。左が 2014 年、右が 2019 年で、公開方針について、必須・推奨・受託・なしを色分けして表示しています。青色がデータ公開ポリシーなしですが、2014 年から 2019 年になると、ほとんどの分野においてデータポリシーなしが減ってきていることが見て取れます。



（図 2）

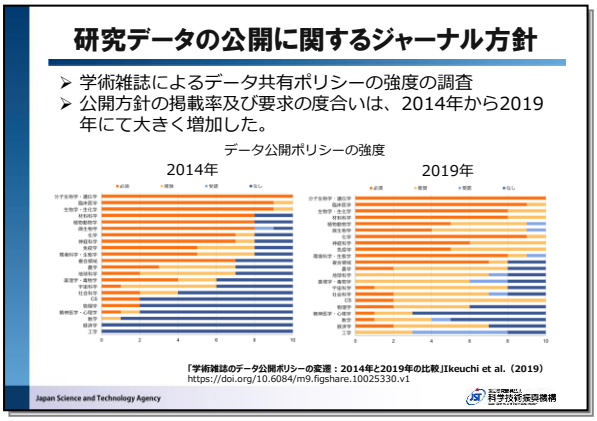
以上のように、研究者が研究成果論文を発表する際、その根拠となるデータ公開を求められる場面が多くなっています。

このような状況を受けて、J-STAGE Data を立ち上げ、J-STAGE の登載記事に関連するデータを登載・公開するデータリポジトリを提供することにしました。その方針としては、全てオープンアクセスで公開し、登載データには DOI を付与するという規格にしました。

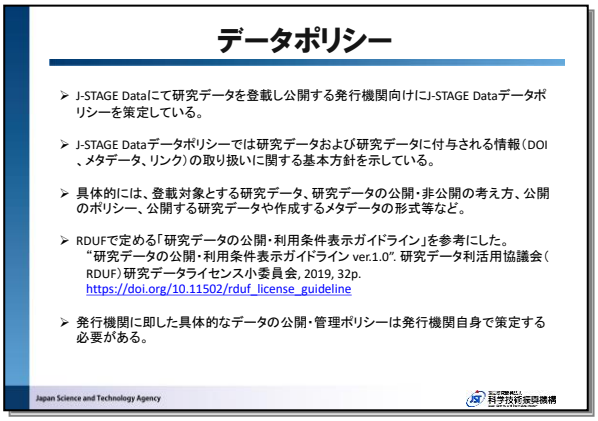
J-STAGE Data の概要

J-STAGE Data は 2020 年 3 月にリリースし、現在はパイロットジャーナルによる試行運用を行っています。登載できるデータは、研究成果論文の根拠となった研究データや、その記事を補足する図表等のデータです。機能としては、J-STAGE 登載の記事との相互リンク、他のデータベースからのアクセス、Google Dataset Search や Dimensions からの検索が可能です。

また、J-STAGE Data を立ち上げるに当たり、推奨するデータポリシーを策定しました（図 4）。簡単に言うと、研究データや研究データに付与される情報（DOI、メタデータ等）の取り扱いに関する基本方針です。もう少し具体的に言うと、J-STAGE Data への登載対象とする研究データや、その研究データの公開・非公開の考え方、公開のポリシー、公開する研究データや作成するメタデータの形式などをポリシーの中で定めています。このポリシーを定めるに当たって



（図 3）



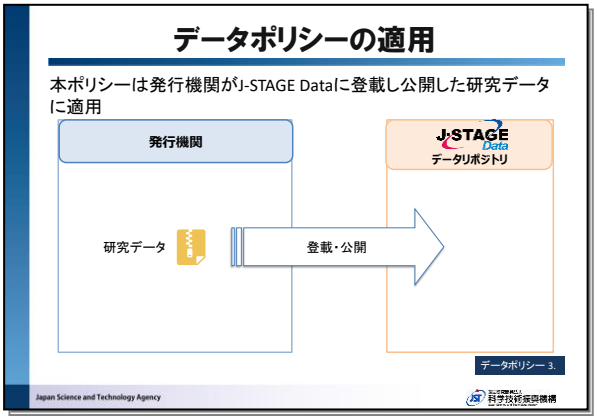
（図 4）

は、研究データ利活用協議会（RDUF）の定める「研究データの公開・利用条件表示ガイドライン」を参考にしています。なお、発行機関に即した具体的なデータの公開・管理ポリシーは発行機関自身が策定する必要があります。

今、発行機関という用語を使いましたが、少し分かりづらいと思うので、データポリシーの中で書かれている状況を整理した図をお示しします（図 5）。データポリシーの適用範囲は、発行機関が J-STAGE Data に登載し、公開した研究データとなっています。この定義は詳しく紹介しませんが、発行機関が研究データを J-STAGE Data に登載することは少し後のフローにも出てくるので、頭に入れていただければと思います。

J-STAGE Data への登載対象となる研究データは、J-STAGE で公開されている、または公開予定の記事・論文にひもづくデータ、すなわち論文記事の根拠となったデータ、あるいは研究の実施過程で得られた情報および関連情報（表、図、音声・動画、ソフトウエア、プロトコル、その他ポスターセッションやプレゼンテーションの資料など）です。

公開・非公開については、著者および発行機関が判断して決定することとしていますが、その判断の際に留意すべきデータ公開への制約があります。研究データに含まれる内容（機密性、プライバシー等）や、研究当事者の要望等によって、データ公開に制約が生じる場合があります。例えば分野・研究コミュニティなどの慣習で公開制限が一般的な場合や、個人情報を含



（図 5）

む場合、国家安全保障等に係る場合、共同研究契約等の契約による制限がある場合、所属機関や研究助成機関などによりポリシーが定められている場合などが関わることがあるので、このようなところに留意しながらデータ公開をしていただくことになります。

なお、これらの制約の中には、データに適切な処理（個人情報の匿名化処理）を施すことや、一定期間が過ぎるとデータを公開することができる場合もあります。

公開のポリシーとしては、原則としてエンバーゴ期間を設定することは推奨していませんが、何かの事情でやむを得えない場合は、研究データに一定期間非表示にするためのエンバーゴ期間を設定することができます。エンバーゴ期間中はメタデータのみが表示されます。推奨では、エンバーゴ期間は 12 カ月以内としています。

さらに、研究データを公開しない場合でも、研究データの内容の詳細（データ収集方法やデータの性質、精度等）や問い合わせ先をメタデータに記述してメタデータのみ公開することができます。これは、データが存在したことを証跡として残すことで、将来の研究活動を支えることができると考えているからです。

公開する研究データについては、できるだけ再利用可能なように、特定のアプリケーションに依存しない形式（例：PDF より xls 形式、xls 形式より csv 形式）、あるいは研究コミュニティの中で標準化されたものがあればそれにのっとることで、相互運用性が高まると考えていただければと思います。作成するメタデータについては、データの収集方法やデータの性質等、研究データの内容をできるだけ詳細に記述していただければと思います。これも、語彙等について標準化されたものがあれば、それに従うのがよいと考えています。

続いて、研究データの公開のワークフローについて説明します。図 6 の一番左は、J-STAGE に論文を投稿してから公開に至るまでの線です。ジャーナルを発行する機関が原稿を受け付けて、査読・編集した上で J-STAGE への登載作業を行います。図の右側は、研究

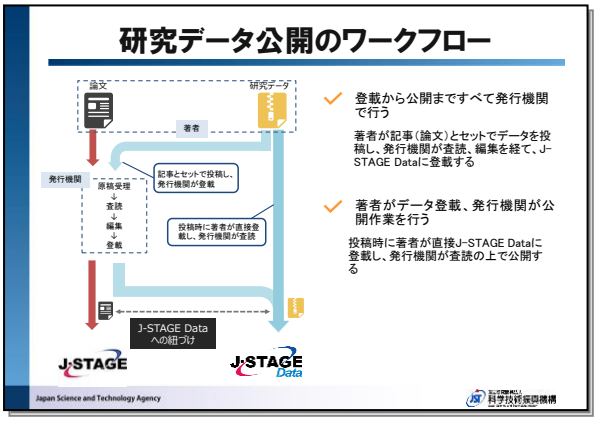
データを登載していく流れです。矢印が 2 本あるのは、2 パターンあると考えているからです。左側の青い線は、著者が記事とセットで投稿し、発行機関が登載するという流れです。こちらは著者が発行機関に研究データを渡し、発行機関が査読・編集を行った上で、J-STAGE Data への登載から公開まで行うものです。右側の青い線は、著者が J-STAGE Data へ投稿し、その上で発行機関が査読して公開作業を行うという流れです。このように、研究データの公開には大きく分けて二つのパターンがあると考えています。

実際に研究データを登載するときの画面を紹介します。図 7 は登載するデータファイルをアップロードするための選択画面です。メタデータ（タイトル、アイテムタイプ、著者名等）についてはここで記載します。ライセンスについては、CC ライセンスを基本として、記載のものから選択するようになっています。

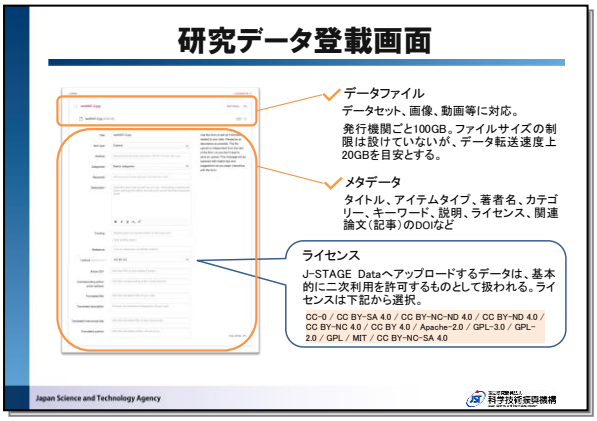
ここまでがデータを登載する機能についての紹介です。ここから先は、データを閲覧する画面等の紹介をしていきたいと思います。

まず、J-STAGE Data のトップ画面を開くと、幾つかの最新の登載情報等がサムネイルと共に表示されます。サムネイルをクリックすると、図 8 のようなデータ画面にいきます。ダウンロードと書いてあるところからデータのファイルをダウンロードします。ダウンロードの左にある「Cite」と書いてある赤いボタンをクリックすると、このデータを引用する際の引用記述を見ることができます。この中に DOI も記載されています。その下にタイトルや著者名、その他メタデータ等が表示されています。右側には閲覧数やダウンロード数、被引用数、データに言及している記事へのリンク、キーワード、二次利用ライセンスの情報等が表示されます。

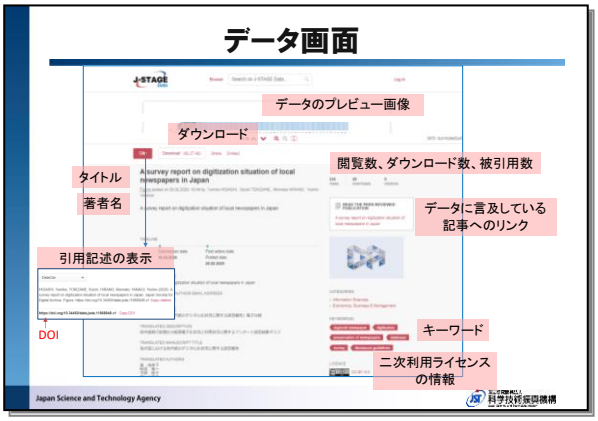
このようなデータ画面と J-STAGE の相互リンクの状況について説明します。図 9 の左側が J-STAGE の



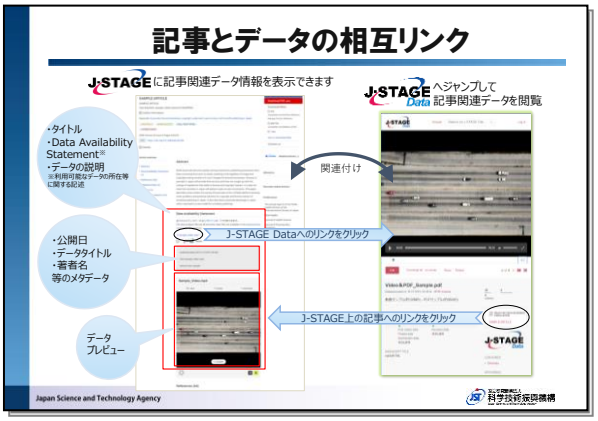
（図 6）



（図 7）



（図 8）



（図 9）

画面です。J-STAGE の記事情報の真ん中あたりにアブストラクトがあり、その下に J-STAGE Data のデータ情報が掲載されます。タイトルや、Data Availability Statement といって、利用可能なデータの所在等に関する記述を掲載することができます。また、データの説明等についてもここに記載可能です。公開日、データタイトル、著者名等のメタデータが表示された後に、データプレビューも表示されます。

J-STAGE Data へのリンクをクリックすると、該当する J-STAGE Data のデータ画面に飛ぶことができます。J-STAGE 上の記事へのリンクはデータ画面の右側に表示されます。

さらに、Google Data Search や Dimensions から J-STAGE Data にリンクができますが、これらは全てのデータではなく、データ種別によります。

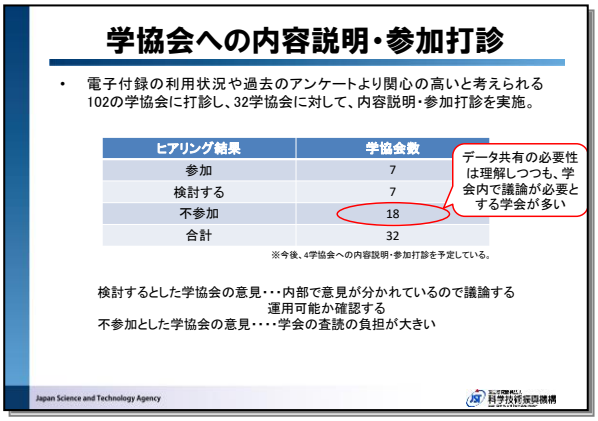
パイロット運用の結果

続いて、パイロット運用の結果を紹介します。パイロット運用期間は 2020 年 3 月から始まっており、2021 年 3 月までを考えています。本格運用に先立ち、必要となる運用フローや登載ガイドラインの策定、事業運営等の検討を行うためにパイロット運用を行い、10 誌程度のパイロット参加を受け付けています。

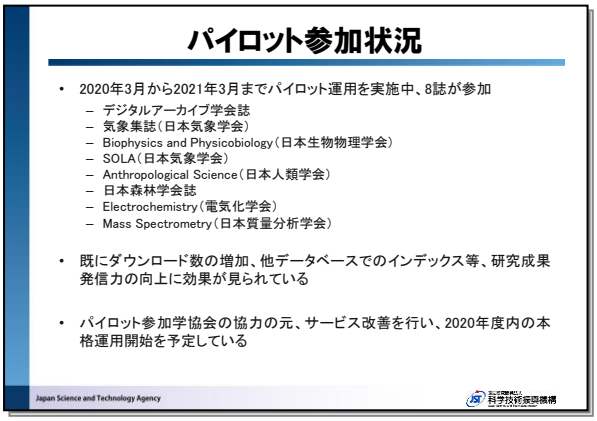
図 10 は現在のパイロット参加状況です。8 誌が参加しており、既にダウンロード数の増加や、他のデータベースでのインデックス等、研究成果発信力の向上が見られています。ダウンロード数の増加等についてはこの後紹介します。パイロット参加学協会の下、サービス改善を行い、2020 年度内の本格運用開始を予定しています。

パイロット運用期間中に学協会に J-STAGE Data の説明や参加打診を行ったので、その状況について紹介します（図 11）。電子付録の利用状況や過去のアンケートから、関心の高いと思われる 102 の学教会に打診し、話を聞いてもいいという回答を頂いた 32 学協会に対し、内容説明、参加打診を実施しました。その結果、「参加」が 7 学協会、「検討する」が 7 学協会、「不参加」が 18 学協会となりました。「検討する」という学協会については、データ共有の必要性は理解しているが、学会内で議論が分かれており、検討に時間がかかりそうだということです。「不参加」という学協会からは、学会の査読の負担が大きいという意見が寄せられました。

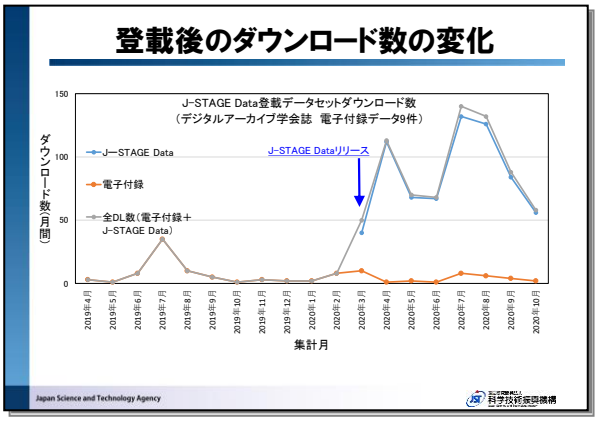
図 12 のグラフは、登載後のダウンロード数の変化



（図 11）



（図 10）



（図 12）

を表したものです。電子付録のときと比べて、J-STAGE Data に登載するとダウンロード数が増加していることが分かると思います。

参考：第 3 回 J-STAGE セミナー

2021 年 3 月 1 日に J-STAGE セミナーを行い、「ジャーナルから見た研究データ」と称して、参加学協会の事例や、国立研究開発法人における研究データの公開の取り組みなどを紹介する予定です。J-STAGE のイベント情報から申し込みができます。今年度は通年で研究データについて取り上げています。本日の内容の補足にもなると思うので、ご覧いただければと思います。

●**八塚**　早速、二つ質問を頂いています。一つ目は、「データを公開する際に、個人情報や機密が含まれたデータについては制約を課すというご説明がありましたが、こういった機密等が含まれていて完全公開できないようなものは、研究者が判断するのか、それとも J-STAGE 側でチェックしていただけるのか、どちらになるのでしょうか」という質問です。

●**加藤**　研究者側で判断していただきます。J-STAGE の登載時に、J-STAGE Data 側でのチェックはかけていません。

●**八塚**　あくまでも研究者側で判断していただく。従って、その責任も研究者側に課せられるということですね。

●**加藤**　そのとおりです。

●**八塚**　次の質問です。「メタデータは、あらかじめ用意されている項目のみ利用できるのでしょうか。あるいは、雑誌や学協会別にオリジナルのデータ項目も用意することが可能なのでしょうか」。

●**加藤**　設定項目については、こちら側で用意した標準的なメタデータ項目を使っていただきます。項目を加えることはできません。

●**八塚**　あらかじめ用意したメタデータ項目のみということですか。

●**加藤**　はい。自由に項目を増やせるということではありません。

●**八塚**　続いて、「先ほどのご説明の中で、著者、研究者自身が J-STAGE Data にアップロードできるようなフローがありましたが、その中で、例えば論文がリジェクトになったのにデータが載ってしまって具合が悪いという場合には削除できるのでしょうか」という質問です。

●**加藤**　基本的に、著者が直接アップロードした場合についても、最終的な公開承認はジャーナル側、発行機関が行うことになっているので、論文がリジェクトされることが想定されていれば公開しないのではないかという意味で、起こりにくいケースだと思います。（後日注：当日パネルディスカッションで補足しましたように、リジェクトの場合、発行機関の中で査読審査をしている中でリジェクトになっていますので、データ自体、まだ公開されていません。また、メタデータも公開されていない状態ですので、その状況の中で削除するかどうかは発行機関が決めることができます。）
なお、公開されたデータについては、できるだけ保持することが望ましいですが、削除になったときについてもポリシーの中で取り決めがあり、削除したときに削除した旨を設けることで、研究データ自体を非公開扱いにすることはできると記載しています。

●**八塚**　次は、「データファイルのサイズは発行機関ごとに 100 ギガとありましたが、それを超えたファイルサイズのデータは扱えないのでしょうか」という質問です。

●**加藤**　パイロット運用期間中は 100 ギガバイトを目安として進めています。100 ギガバイトでは少ないという声も学協会から頂いているので、対応を考えていきたいと思いますが、現状は 100 ギガバイトとしています。

●**八塚**　続いての質問です。「ダウンロード数が変化したのは、学協会の参加の呼び掛けで皆さんがアクセスした、あるいは関係者の皆さんがアクセスしたといったものが含まれていませんか」。最後のグラフのご説明のところだと思います。

●**加藤**　それは多分にあると思います。詳しく分析していませんが、初速があったことは否定できません。

●**八塚**　次に、「既に J-STAGE に掲載されている論文について、後付けでデータを登録することはできますか」という質問を頂いています。

●**加藤**　はい。できます。実績もあるそうです。

●**八塚**　次は、「論文自体がもし J-STAGE に登録されない場合は、データも削除されてしまうのでしょうか」という質問です。つまり、データと論文を一緒に登録したけれども、後で論文がリジェクトされてしまった場合に、データも一緒に消されてしまうのですかという質問です。

●**加藤**　リジェクトされたら、消されることは消されると思いますが、DOI を付与されているので、メタデータ自体の公開は継続されることになると思います。

●**八塚**　メタデータは継続するけれども、データ自体はもうないということですか。

●**加藤**　はい。非公開と同じような扱いです。データ自体はないけれども、こういうデータが過去に存在したという記載と共にメタデータは残ることになります。

●**八塚**　次の質問です。「同じ論文の中で、公開可能なデータと非公開のデータが混在する場合があると思いますが、非公開データも含めて管理できますか」。

●**加藤**　はい。できます。それぞれに別個の DOI を付けられるかどうかという質問でしょうか。

●**八塚**　そこははっきりしませんが、非公開であっても DOI は付けられるということですか。

●**加藤**　はい。そのとおりです。

●**八塚**　分かりました。

次の質問は、「例えば学会や研究会の中で J-STAGE Data を使って制限共有することは可能でしょうか」というものです。

●**加藤**　制限共有はできません。

●**八塚**　一般に公開することが前提ということですね。

●**加藤**　はい。公開となった場合は一般公開です。公開前の登載の中で、学協会の事務局での確認ということは想定しているので、編集委員会等が見るという使い方は想定していますが、もう少し大きな集団、例えば学会内だけの限定公開はできません。