

# 第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

## データ流通市場におけるデータ公開の事例・論点

上島 邦彦

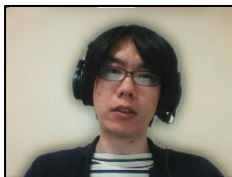
(株式会社日本データ取引所)

### 講演要旨



民間企業のデータ公開にはしばしば多くの障壁が生じます。その背景には法的課題や技術的問題だけでなく、ブランド毀損やビジネス機会の損失リスク、ITリテラシー格差といった諸問題もつきものです。利用規約の作成、業務フローの標準化、提供内容・方式の棲み分け、UI/UXデザインによる事故防止など、さまざまな工夫が行われています。

本報告では、スタートアップ企業の視点から、日本企業のデータ取引によくある論点と、データ流通市場サービスを介してデータを限定公開する事例を紹介します。



### 上島 邦彦

株式会社日本データ取引所に創業メンバーとして参加。事業企画・プロダクト開発全般に従事。企業間のデータ調達・販売支援を始めとして、データ製品開発や社内データ戦略企画などにも携わる。研究データライセンス小委員会メンバー、データ流通推進協議会技術基準検討委員会副主査。

### 日本データ取引所 (J-dex) 会社概要

弊社はその名のとおおり、データの取引を専門に行う会社として創業し、今年で5年目に入ります。企業と企業の間を取り持つことが多いです。データ活用戦略の立案やデータ調達の仲介——本日の論点でもある、どこまで公開・共有できるかも含めた企画・設計に携わっています。

利用条件やプライバシーポリシー、セキュリティポリシーの整備など、組織間でデータを流通させるときの法的相互運用性の向上についても、弁護士と連携しながら実務支援をさせていただいています。

研究データ利活用協議会のライセンス小委員会にも参加している他、データカタログやデータ品質に関するガイドラインの策定、国際標準化活動にも、微力ながら貢献しています。

本日は、企業の一従業員の目線でお話をさせていただけます。企業がデータを公開・共有するときの肌感覚——ソフトローがどのようなものであるかを実感していただければと思います。キーワードは、「鍵のかかった部屋」をいかに持ち運ぶかです。

### データとは何か

データとは、ざっくりと言うと記録された事実であり、価値ある情報の素材です (図 1)。日本工業規格が 1994 年に、情報とは何か、データとは何かを定義していて、これがよく参照されます。情報がデータとして存在を認められ、利用可能となるまでに、複数の工程が必要であることは皆さんも実感的に理解されていることでしょう。多くの工程を経てデータが流通しています (図 2)。ビジネスや学問、行政を支えるデ

ータが、世界中で動き回っています。他の製品やコンテンツと同じで、何かしら対象物に関する記録を行い、それを蓄積、転送して統合する、いわゆる原料生産に当たるプロセスがあり、外に出してもいい製品に加工するプロセスを経て、使う側が評価したり、それを基に判断・行動するという一連の流れを回す。これを広く「データ流通」といっています。

「データ製品」という用語もあります。先ほど申し上げた、概念としてのデータではなく、外に出せる、売ってもいい、他の方にお使いいただける、生の情報を加工したデジタルコンテンツの総称です。その製品情報を要約・整理したものを、「データ製品カタログ」と呼んでいます。この言葉が指す対象は、学術研究におけるデータ出版の対象データや、研究データリポジトリとほぼ同じかと思えます。

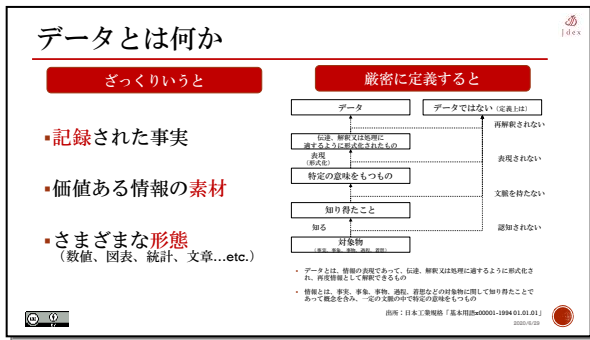
データ取引についても簡単に説明します。ごく抽象的に考えると、誰かが持つデータを他の相手に見せたり渡したりすることは、広く「データ取引」だと言えます。取引交渉によってお金とデータを交換するだけでなく、お互いにデータを開示したり、共同研究のよ

うに、データ共有をして共同で論文を書くことも含まれます。オープンサイエンスのために世の中にデータを公開することも、対価のない取引だと考えられます。

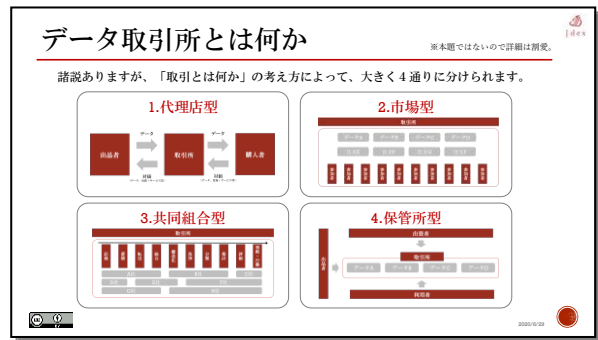
弊社名にもある「データ取引所」には、いろいろなパターンがあります(図3)。単純に仲介する「代理店型」。注文と商品をマッチングする「市場型」。データ流通プロセスの一個一個を皆さんで担う「共同組合理型」——これは魚の市場に近くて、同じ場所に皆で集まってやる方が楽し早いというもの。そして、研究データリポジトリはこちらに近いと思うのですが、取引所が出品者からデータをお預かりし、そこに第三者が出資し、また第三者の方が利用する「保管所型」。これらをひっくるめて「データ取引所」と呼んでいます。

### データ公開の方法整理

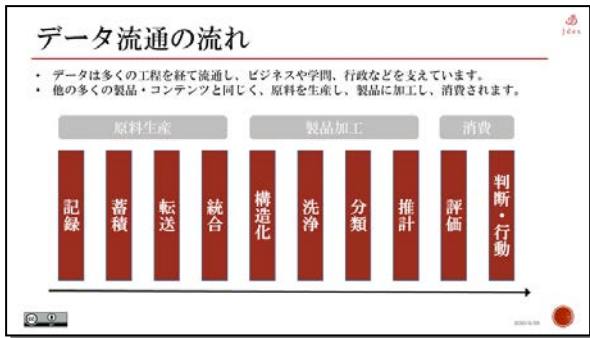
データの公開・共有にグラデーションがあることは、企業のデータ取扱担当者も日々感じながら仕事をされています。例えば、図4の左下——企業のデータ取扱担当者が何かデータを持っているとして、それを誰に



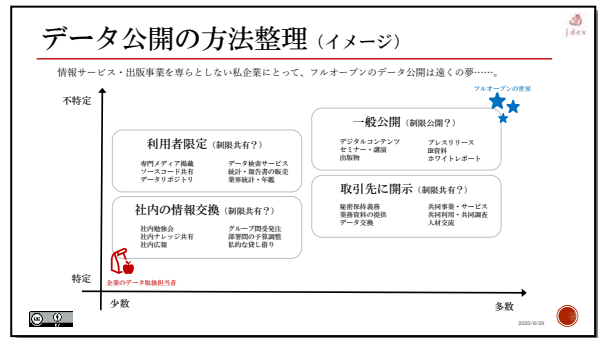
(図1)



(図3)



(図2)



(図4)

も公開しないことはよほどあり得ず、少なくとも社内の誰かとは共有するでしょう。そこから出発して、右上にあるように、条件指定や利用制約なしで、フルオープンで公開するという大きな目標に向かっていく。

2つの軸で考えてみます。横軸は「どれくらい多くの方にご覧いただくか」（多数・少数）。縦軸は「どれくらい顔の見える方にご覧いただくか」（特定・不特定）。この間にグラデーションがあり、どこまで多数の人に、どこまで不特定の人に情報を出すかによって、情報を出す形式・方法が変わると考えてみます。

例えば、社内の情報交換のために制限共有する（左下）。利用者を限定して、ある程度専門知識や業界固有の事情が分かる方に公開する（左上）。あるいは取引先と契約を結んだり、共同の事業・サービス・プロジェクトを組んで開示する（右下）。私ども（民間企業）が「一般公開」と呼ぶものは、実質的には「制限公開」ですが、デジタルコンテンツとして世に問うたり、セミナー・講演でグラフを見せたり、出版物を出したりするときには、著作権をはじめ各種の制約が生じます。もろもろの法的観点の整理なども面倒になってきます。

## 企業のデータ公開にありがちな障壁

企業のデータ公開に当たって、ありがちな障壁が5つあります（図5-6）。大きいのはやはり法的課題です。一番大きい国際法のレベルから、国内の法令レベル、企業間での契約レベル、そしてここが今一番やっかいなところかと思いますが、実務レベル。それぞれで問題が生じます。データ作成者が——研究データでいう

と研究者が——これら全ての知識を持つことは、かなり難しいのではないかと感じています。

技術的問題もあります。単純に言えば、知識です。データに関する知識と業界に関する知識の両方が分かる人は、昔も今も将来も少数派でい続けるでしょう。また、データ製品開発のためにいろいろな技術が考案され、提供されていますが、慎重を期すれば期するほど手間や労力がかかります。「そこまでして外に出したいデータなのか」と、皆さん悩みながら作業されています。それから、先週から今週にかけて各地で事故が起きたとも報じられていますが、安全にデータを移転するためのセキュリティ、プライバシー、証跡管理などもよく問題になります。これらはハードな制約です。

ソフトな制約もあります。よく聞くのは、企業・部署としてのブランドが毀損されないかといった何となくの不安感や、ステークホルダーの意向や考えの板挟みになってなかなか身動きが取れないということです。

ビジネス機会を逸するのではないかという不安も、現場の方や経営者の方が感じることは多いようです。そもそも「データを持っています」と公に言わないこと自体が競争力になりますし、データの公開・共有が自社のブランディングにつながる、そういうことが好きな人材が来てくれる、その作業自体で知識習得できると分かっているにしても、投資対効果といえますか、どれくらい労力をかけていいのかよく分からないという声はよく聞きます。個人の感覚ですが、家に本がどんどん積み上がって減っていかないことに似ていますね。ITリテラシー格差などもよく問題視されます。

### 企業のデータ公開にありがちな障壁 (1/2)

**法的課題**  
 国際レベル：通商協定、越境データ移転、ローカリゼーション規制  
 法令レベル：営業秘密、知的財産権、個人情報、パブリシティ権  
 企業レベル：秘密保持契約、データ利用規約、情報管理指針、社内決裁  
 実務レベル：利用条件の整理、契約書の起草、ライセンスの設定

**技術的問題**  
 統計調査・データ分析と専門分野の双方に精通した人材は少数派（昔も今も将来も）  
 「欲しいデータ」「外に出していいデータ」の判断も意外にむずかしい

データ製品開発には多様な技術が介在する（慎重を期すほど高コスト体質に）  
 デバイス性能理解、データ品質評価、データ洗浄手順、匿名化・曖昧化

安全なデータ移転のためのセキュリティ確保、プライバシー保護、証跡管理  
 業務フロー自体に根ざすリスクを減らすには、働き方から変える必要がある

(図 5)

### 企業のデータ公開にありがちな障壁 (2/2)

**ブランド毀損**  
 業界慣習、社内外の視線、敏感な専門家、無関心な一般生活者の板挟み  
 表上リスクに当意即妙で応じられる広報スキルと法務リソースを持つ担当者は少数

**ビジネス機会の損失リスク**  
 「データの堅持」を従えること自体が競争力になる  
 サンプルデータや項目定義だけでも技術流出リスクが懸念  
 サプライチェーンの上流・下流で、交渉材料として使われる情報開示

得られる利益（ブランディング、人材獲得、知識習得）の不確実性  
 顧客層・現場ともに「すべきた」と分かっているでも、設備・人員の投資判断を行わずに  
 自社の積ん置きがどういっけか一向に減らないことに似ている

**ITリテラシー格差**  
 組織内のデータ活用スキルのばらつき、データ活用部門の大いなる忙殺  
 取引相手との知識ギャップ（スーツとギークの、理論と応用の、大企業とベンチャーの...）

(図 6)

## 企業のデータ≒「鍵のかかった部屋」

つまり、企業のデータは「鍵のかかった部屋」のようなもので、中に何かがあり、価値もあるのだからけれども、その価値を知るには鍵を開ける必要があって、なかなか入りづらく、中にいる人、外にいる人、双方にとって苦しい状況に陥りがちです。

そんな中で、誰もが安心して自由にデータを使える——先ほどの図の一番右上までたどり着くにはどうすればいいか。一般論でいうと、「鍵のかかった部屋」のドアをロックして中に招き入れてもらい、その部屋の住人になればいい。鍵を手に入れて、長居して、ゆくゆくはゴールを目指すといったことを考える。

少し常識から外れて、概念的に考えると、そもそも鍵があるからいけないのだ、出入り自由にすればいい。部屋そのものを解体して、壁をなくせばいいといった、空間デザインのような発想もできるでしょう。そして、私が最近考えていて、本日も説明したいのは、「部屋ごと持ち運ぶ」こともできるのではないかな。

「鍵のかかった部屋」をどう扱うか。企業データの取扱法になぞらえると、20世紀的方法として、組織のメンバーシップを入手する、データのオーナーシップを持つ、データの利用権限を付与してもらうなどがあります。これらはデータポリシーの整備など、制度・法務面の対応でクリアできます。加えて昨今増えているのは、コミュニティを運営して、そこで公開・共有していくことです。いわば21世紀的方法です。一定のライセンスを設定して、より多くの方が使えるようにする。さらには、弊社がまさにそうですが、機能的なサポートを受ける、マーケットプレイスに出品するといった挑戦も出てきています。大きく2つ、説明します。

まず、データの権限を付与してもらう事例です。先ほど申し上げた4つの課題に対して、リーガル面だけではなく、サービス面、システム面、コンテンツ面でも工夫することで、総体としてリスク・コストを下げる手法が取られます。利用規約を作成したり、業務フローを標準化して技術的な問題が起きにくくしたり、

システム側で事故が起きにくくしたり、出せるものと出せないものを切り分けたりします。リテラシーに関しては、もう頑張るしかないという感じですが。

弊社が関わった事例では、しょうゆや豆乳のメーカーとして知られるキッコーマンと、PR会社のオズマピーアール、東京大学大澤幸生研究室の4者による共同研究があります。約20社間でデータ流通の個別の利用許諾を得て、データ活用の枠組みを作りました。これは、「やればできるけれども大変だ」という事例です。

2つ目は、データ製品の出品登録・購入リクエストの事例です。(利用許諾の手続きを)システム化してしまうことです。私どもの挑戦として、データの出品と購入ができる会員制のマーケットプレイス「JDEX™」を11月ごろから始めようとしています(図7)。データの制限公開と制限共有を線引きするための微妙な取り決めなど、細かいところをサポートする機能を備えて、実際にお使いいただこうと思っています。データのカatalogを作れるようにして、データ自体を出品しなくていいようにしたり、審査制にしたり、特定の相手だけに注文・出品できるようにしたり。匿名加工やライセンス整理など、実務面のサポートもシステム側で行おうとしています。使い慣れれば簡単にできると期待していますが、もちろん日本ではまだ普及していません。「まだまだこれから」の事例です。



(図7)

## 論点のまとめ：「鍵のかかった部屋」をいかに持ち運ぶか

企業のデータは、なかなか鍵を開けて外に持ち出せない。その状況で、「鍵のかかった部屋」をどう持ち運ぶか。境界線をずらす、動かすという発想をした方が、物事が進みやすいのではないかと思います。

まとめると大きく 3 つの論点があります (図 8)。まず、不特定多数と特定少数の間に多様なデータ公開の仕方があり、ソフトローによって制約もそれぞれです。また、制限公開あるいは制限共有が選ばれる理由は、複合的で一様ではなく、トータルでリスクとコストを下げる必要があります。1 つの解決策として、実務的に頑張って労力を割く従来の方法も有効ですが、今後は、ルールやシステムで省力できるといいのではないかと考えています。

論点まとめ

1. 不特定多数と特定少数の「間」に多様なデータ製品の形態  
暗黙知や社会的慣習のちがいによって、制約もそれぞれ
2. 制限公開が選ばれる理由は複合的で一様でない  
法的課題、技術的問題、ブランド毀損・ビジネス機会損失リスクなど
3. 従来手法も通用するが、ルール&システムで省力したい  
「身内になる」と「一般開放（制限付き）」の2択以外の道が開ける

(図 8)