

ニューズレター

別冊

全国公共図書館協議会

平成30年1月30日

(〒106-8575 東京都港区南麻布5-7-13 東京都立中央図書館内)

【全国公共図書館協議会研究集会講演記録】

テーマ 図書館とオープンデータ

講師 国立情報学研究所 准教授 大向 一輝

平成29年6月30日(金)に開催された全国公共図書館協議会研究集会の講演記録を別冊としてまとめました。

今回は、大向准教授にオープンデータの基礎をなす制度面及び技術面の再利用性や図書館とオープンデータの関わりについて御講演いただきました。

なお、この講演記録は実際の講演内容を再構成したものです。

自己紹介

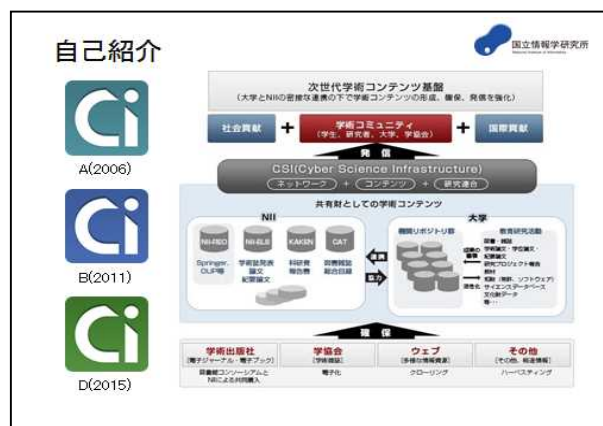
国立情報学研究所から参りました、大向と申します。今日はどうぞよろしくお願い致します。今日は、図書館とオープンデータというタイトルで、ご依頼をいただきましたので、最近の情報などをこれからご説明したいと思っております。



〈写真1〉研究集会の様子

最初にご紹介をさせていただきたいのですが、私は大学の図書館にかかわる仕事を主にやっております。公共の図書館でもお使いか

とは思いますが、論文の検索ができる CiNii や、大学の図書館の総合目録を検索できる Ci Nii Books、博士論文を検索できる CiNii Dissertations、これら3つの CiNii シリーズを、主に面倒を見るという仕事をしております(図1)。



〈図1〉自己紹介

国立情報学研究所というところは、古くは学術情報センターと呼ばれていて、NACSIS という名前でしたが、それが2000年になってN

II、国立情報学研究所という名前に変わりました。もともとのミッションが、大学の図書館と協働して総合目録をつくっていくというのですが、大学の総合目録というのはかなり特殊なシステムになっています。実際、公共図書館ですと、書誌データはそれぞれの自館のところへ MARC 購入という形で OPAC に登録されて公開されていると思います。

今、全国に 700 から 800 ぐらい大学がありますが、ほとんどの大学は、一たん我々が管理しているデータベースに何の本を持っているかという情報を登録しなければいけないという仕組みになっています。学研究において手に入らない本があったときに諦めるわけにはいきませんので、どこの図書館に行けば借りられるのか、相互貸借のシステムが全国に整備されておりますので、そのためにも情報を 1 箇所に集めなければならないという目的で、NACSIS という組織がつくられました。設立から 30 年以上経過していますが、データベースをメンテナンスし続けているという組織です。

それが 2000 年になり、情報学の研究の中心にもなりなさいというミッションが加わりました。ある面では大学の情報工学科などとほとんど同じような主義、志向をしている研究者もいますし、一方で大学の図書館のインフラという意味での情報サービスをやるという 2 つのミッションを抱えた組織であります。

もともとは図書館の紙の本をいかに管理するかというところだったのですが、特に大学ですと、最近本を買うのはどんどん減っています。電子ジャーナルを買わなければいけない。論文の契約更新をどんどんしなければいけないのですが、今、値上げがとどまるところを知らないで、皆非常に悩んでいるところです。そうすると少しでも安くならないかと大学が組んで、出版社に交渉するというのをやったりする仕事もあります。

あと最近、機関リポジトリと申しまして、大学に所属する研究者が出した成果を、自分

の大学から発信するというのを強く求められています。そのためには、サーバーを全部自分たちで用意していかなければいけないのですが、これを支援するために、NII のほうでクラウドサービスを用意しています。そこに参加すると、サーバーの管理は我々が一手に引き受けます。そこで載せる論文や成果だけをどんどん登録してもらって、ネットを通じて公開する。そんなこともやっていたりします。

あとは、研究だと研究費というのを、個々の研究者はいただかなければいけないのですが、それをもらったからには報告書を書かなければいけない。その報告書を取りまとめてネットで公開し、納税者の皆さん、あるいはほかの研究者の皆様にきちんと成果をお伝えするためのデータベースなどをつくってたりします。

ということで、今 10 個ぐらいサービスをやっています。ではその 10 個のサービスを何人の職員でやっているのかというと、10 人ぐらいでやっているの、極めて大変ではあります。もちろんさまざまな大学の職員を含めて、たくさんの人の支援をいただきながら、そういった情報サービスを全国区で展開していくというのが、NII という場所のミッションです。一部のサービスをお使いの方もいらっしゃるかもしれませんが、実はそういうところであるというところだけ、最初にご紹介させていただきます。

私自身は情報系の研究者です。ソーシャルメディアのブログ、ツイッター、フェイスブックなど、人々がネットを通じてコミュニケーションをするということが、2000 年初頭からすごい勢いで盛り上がっています。そこに対してどんなコミュニケーションが行われているのか、どんな仕組みやソフトを提供すると、より議論が活性化するのかといったことを研究テーマとして博士号まで取りました。

その後就職したのが先ほどの NACSIS の総合目録システムなどをやっているところです

が、実は全然そんなことは知らなくて、「ああ、うちで CiNii やっているんだ」みたいな非常にのんきな感じで見ていました。当時はかわっていませんでしたが、ネットサービスとしての CiNii はあまり出来が良くないと思っていたので、「例えばグーグルは検索ボックスが 1 個しかないのに、何で OPAC は画面を開くと 5 個も検索ボックスがあるんですか」と素朴に聞いたり、「1 個でいいんじゃないですか」と言う。すると「それはこういう理由で」との説明を受けるわけですが、ネットの一般ユーザーからするとこれはちょっと受け入れられないなど、いろいろ議論をしているうちに、「そんなに文句を言うのなら、おまえが 1 回つくってみろ」と言われ、そこで手を上げてしまったが最後、そこから約 10 年間 CiNii を担当しています。

その中で、学術情報流通に関しても学ぶことが非常にたくさんありましたし、そこで自分の Web を対象とした研究の成果を少しでも入れられないかなと思いながら、今も CiNii 等々のサービスにかかわっています。

Web に関しては幾つか著書もごさいますので、もし機会がありましたら、手にとっていただければと思います。

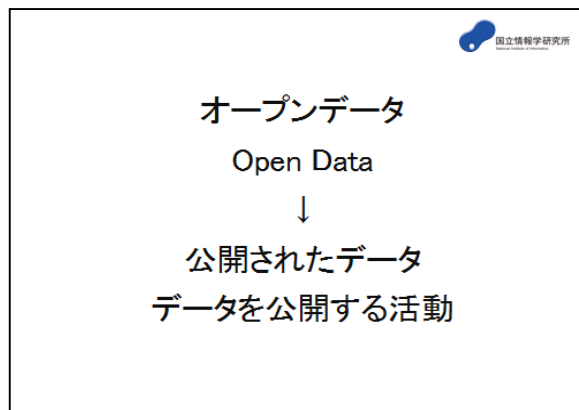
そういった学術情報に係る研究であったり実務をやる傍ら、Web の世界ではどんどん進化が進んでいるというのも実際のところです。特に Web という場所が、単に人がネットにアクセスし情報を自分で読んで理解するということを超えて、データそのものを共有する場所に少しずつ進歩しているという現実があります。それも少しご紹介したいと思いますが、データを皆でシェアするための Web というところから、今日のテーマにもなりますオープンデータという流れが出てきているということです。

オープンデータの詳細をこれからゆっくりと説明していきたいと思います。特に先ほどご紹介にもありましたが、政府や地方自治体が非常に急ピッチでデータの公開をやっている

ます。それも最初からどんどん出てきているわけではなくて、2010 年ぐらいから徐々に、これは一体何なんだという議論から始まり、少しずつ皆さん経験を積まれて今に至っているわけです。その中で、特に私は、技術面からどういったことをしたらいいですよというアドバイスをしてまいりました。

今は、内閣官房のところでデータワーキンググループという集まりがあるので、そこに行ったり、あとは総務省を中心として一般社団法人ができていますが、そこで委員会をやったり、あとはオープンデータの技術的な面をもっと広めていこうというので NPO をつくったりといったことをやってきております。

今日はそういういろいろな図書館の仕事をしている立場や、オープンデータの普及もしたいと思っている立場から 2 つをくっつけたようなテーマをせっかくいただきましたので、そのことについてお話を進めていきたいと思っています。



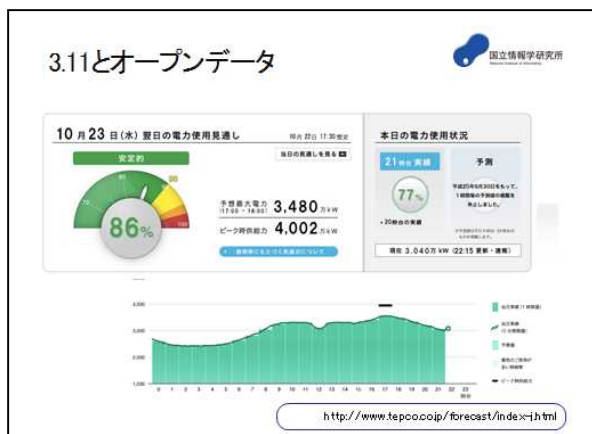
〈図 2〉オープンデータとは

オープンデータとは何ですかというところですが、2 つの英語の単語が並んでいる。open、data という、公開されたデータのことをオープンデータというのでしょうか。あるいは、データを公開する活動のことをオープンデータと呼ぶのだろうと〈図 2〉。もうこの 2 単語を見ると、それ以上言いようがないので、それだけ理解していただければ、今日のお話は終わりと言ってもいいのですが、意外と七、

八年こういう議論をしていると、専門的にもあるいは一般の方々も含めて議論が深まってきたいて、単にデータを公開するだけではない、単にネットにぽんとデータを置いておけば、それがイコールオープンデータではないというのが、少しずつ今のコンセンサスとなっています。ですので、ある意味専門用語化していて、ちゃんとその意味を理解していただいた上で、実際にオープンデータ活動をしていただきたいという思いもあります。

データを公開するというのが、まずいいことなのかどうなのかという経験を我々が最初に非常に強く経験したのが、2011 年の 3.11 のときであります。

当時のことを思い返していただければ、地震があつて津波があつて原発の問題があつてと、非常に世の中が混乱していました。私は東京に住んでいますので、地震そのものでものすごく大きな被害があつたわけではありませんが、一番大きな問題は電力であつたわけです。最大何ワットの発電能力があるうちの何%を使っている、これ以上使うと首都圏が全部ブラックアウトしてしまいますというニュースがあつたと思います。その当時、東京電力あるいは関西電力を含め電力会社は、実際の発電能力と実際にどれぐらい使われているかということを、きちんと情報公開していたわけです。



〈図 3〉 3. 11 とオープンデータ

これは、震災の後、何カ月後かの画面だったと思いますが、こうやって非常に見やすい

形で 86%使っていますと出ていたわけですから〈図 3〉。グラフも出ていて、どういう推移だったのか、非常に親切かつ丁寧に情報公開をしていると思います。一方で、電力のことがいかに気になるからといって、3.11 直後のありとあらゆるものが混乱している中で、1 時間に 1 回、ネットにこちらからアクセスしに行くということをした方が、どれだけのものか。ほかのこともたくさんある中で、チェックを能動的にしなければいけないというのはなかなか大変なことですし、それをこまめにやっていた人というのは実際にいらっしゃると思います。



〈図 4〉 3. 11 とオープンデータ

非常にすごいなと思ったのは、そんな中でも、いろいろなエンジニアの方々が当時ボランティア的な活動をされていて、このグラフから数値を読み取るプログラムをつくって、今は 3 千何百万ワットと読み取る。それを例えばメールで配信するとか、ツイッターで伝えるとか、そうやって情報を送りつけてくれる形に無理矢理変換してやっていた。かなり腕の立つプログラマーの方もいらっしゃいましたが、皆が皆そういうスキルを持っているわけではない。これも冷静に考えると、こういうグラフをつくれるデータを持っている電力会社は、もとの数値データも持っているでしょう。であるならば、凝った形で見せるのではなく、数字が並んでいるだけでいいので、それだけを出してほしい。そうすると、先ほどの絵から数字を読み取るスキルはなくても、

数字を使って何か加工してサービスをつくれるという方はいっぱいいらっしゃるのですが、そういった方に対していいのではないかと議論が、震災直後からわき上りまして、現実1週間かそれぐらいのかなり近いタイミングで、電力会社はこのようなファイルを公開することになったわけです〈図4〉。

具体的には、何月何日の何時に何ワット使っています、余力はこれぐらいですというのが、カンマで区切られたテキストでひたすら更新されていく。ただそれだけです。見た目は全然美しくない殺風景な情報ですが、これが手に入るとなった途端、全国のプログラマたちが、では私はiPhone用のアプリをつくるよとか、アンドロイド用のアプリをつくるよとか、ツイッターのボットでは自動的につぶやくプログラムとか、そういったものも非常にたくさんできました。



〈図5〉 3.11とオープンデータ

これはアンドロイド用の電力表示アプリのごく一部です〈図5〉。ちょっと検索するだけでもものすごい数が出てくるのですが、いろいろな見方でアプリを公開して、それを使うと手元で情報がすぐ手に入る状況をつくり上げられたりしたわけです。

もちろん平時であれば電力会社のサイトのような見せ方も丁寧でいいと思いますが、そういうことだけではなくて、元データを公開することによって誰かが協力してくれるという体制を、契約等々一切なしに、電力会社と市民の方々との間で実際取り結ぶことができ

た。そういう例ではないかと思います。

そういったことで、データをそのまま公開するというのは結構意味がある。もちろん日本では今の3.11は非常に契機になったわけですが、その前後から世界的にもそういうことをやっていかないかということが、少しずつ議論としては盛り上がってきたという現状があります。

その中で、オープンデータというのを改めて定義していこうという議論も進んでいて、Web上で公開された開放的な、あるいは再利用しやすいデータのことをオープンデータと呼ぼうじゃないかとなってきました〈図6〉。

〈図6〉 オープンデータとは

さらに言えば、開放性とか再利用性とかの定義の中には、制度としての使いやすさと技術面での使いやすさという2面があります。これも細かくは後でご説明したいと思います。そういったものを満たすようなネットで公開されたデータのことをオープンデータと呼ぶのがよいだろうというように議論は整理されてきています。

オープンデータという言葉以外にも、ビッグデータとかパーソナルデータとか、非常に最近世を騒がせる用語もあります。基本的にはデータを使ってビジネスを起こしたり、世の中をよくしていこうという動きはかなり強まっております、そのうちの1つとしてのオープンデータと位置づけることもできるかもしれません。

最初にも申し上げましたとおり、オープン

データというのが特に公共セクターで非常に盛んに勢いを持っている。オープンデータという定義から考えても、別に政府、行政機関だけがやることではないと思うのです。個人がやるというオープンデータもあるでしょうし、企業がやるデータもあると思いますが、特に公共セクターで急速に進んでいます。

政府や行政というのがもっと開いていくことができないかという、オープンガバメントと呼ばれる潮流が議論されていて、その中の実現手段としてデータを公開するということも結構大事だということが言われています。

政府や行政を開くということは何のためにやるのかというのが、また少し議論があって、例えば、透明性を高めるために開いていきましょう、あるいは市民にもっと行政そのものに参加してもらおうであるとか、あるいはそこから新しいビジネスを生み出してもらって経済発展をしていきましょうとか、こういった3点ぐらいに、オープンにしていくことの意義というのは定義づけられています。こういったことがかなり強く出るのが公共セクターということですので、そういった面からも、行政、自治体のデータ公開が進んでいるという面があるのではないかと考えています。

スライドでは、右側に、ある講演会の写真があって、真ん中に立っている人が手を広げて、「Raw Data Now!」、生のデータを今すぐ公開してくれと叫んでいます。これは、TED という世界的にも社会貢献的な活動をしている人が講演をする有名な会議があるのですが、そこである人が話している。この人が誰かという、下の吹き出しでティムと書いてある。ティムというのはティム・バーナーズ＝リーという人です。この名前にぴんときた方もいらっしゃるかもしれませんが、ティム・バーナーズ＝リーというのは、Web を発明したその人であります。

Web の開発者というのは明らかに1人とわかっていて、それがティム・バーナーズ＝リーです。イギリス人ですが、この功績でナイ

トの称号ももらっていて、今はサー・ティム・バーナーズ＝リーと呼ばれています。今 Web の次のステップを考えると、こういう生のデータをどんどん皆で公開していくことによって世界を変えていこうという、旗を振っている張本人でもあったりします。

政府オープンデータの系譜



- ・ 2008 公共セクター情報(PSI)利活用に関するEU指令
- ・ 2006 英国Office of PSI
- ・ 2008 OECDリファレンス宣言・勧告
- ・ 2009 米国Open Government Initiative
- ・ 2009 米国・英国オープンデータカタログサイト
- ・ 2010-2011 フランス・ドイツでの取り組み
- ・ 2012 電子行政オープンデータ戦略
- ・ 2013 G8オープンデータ憲章
- ・ 2013 政府オープンデータカタログサイト
- ・ 2015 政府標準利用規約(第2.0版)
- ・ 2016 官民データ活用推進基本法

〈図7〉政府とオープンデータの系譜

こういった流れの中で、特に政府レベルのオープンデータというのは、着実に進歩してきています〈図7〉。ここ十数年の流れをざっと書いていますが、頭のほうではまだオープンデータという言葉ははっきり出てきていなかったのですが、オープンガバメントをやるためのオープンデータということをかなり強く打ち出したのは、2009年のことです。

このころはオバマ大統領が就任した直後であり、一つの政策のあり方として、オープンガバメント、オープンデータを推進していきましょうということが言われました。

アメリカ、イギリスというのはこういったものを先導する力が非常に強いので、どんどん先行していき、その後ヨーロッパ諸国がついていく形で、大体2010年前後になると、いわゆる大きな先進国と呼ばれるところは、一通りオープンデータについて何らかの指針を持つようになっていきました。

下のほうで色をつけているのが、日本の動きでありまして、2012年に日本でも、政府としてははっきりとオープンデータを政策の中に位置づけるという電子行政オープンデータ戦略というものが打ち出され、それをもとに各

省庁がオープンデータの活動をしていくということが決められていきました。

2013 年にサミットが北アイルランドであったのですが、そこで G 8 の国々は皆オープンデータを進めましょうというオープンデータ憲章にそれぞれサインをしました。その後はサミットがあるごとに進捗状況をお互い確認するという、かなり強いコミットメントを各国が約束することになり、この時点で日本もオープンデータをやるということが国際公約になっていたわけです。

こういったオープンデータ戦略や、オープンデータ憲章というバックグラウンドを経て、政府オープンデータのカatalogサイト、政府標準利用規約というのがつくられ、去年の年末はついに日本においてもオープンデータにかかわる官民データ活用推進基本法が議員立法でできたということです。これも我々にとっては非常に大きなニュースになっています。この後もう少し説明していきたいと思います。

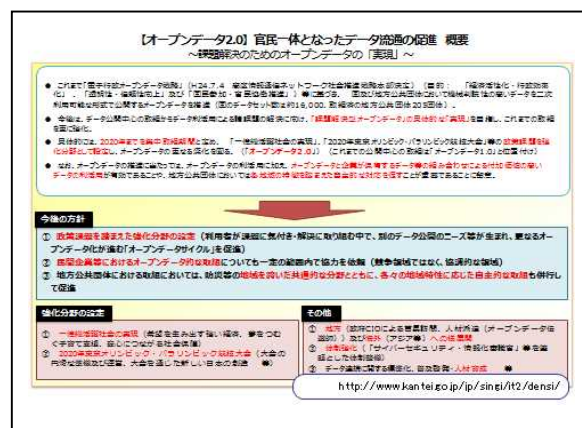


〈図 8〉

このスライドの下部にアドレスが書いてありますが、ホワイトハウスのページです〈図 8〉。whitehouse.gov/open というすごくわかりやすいアドレスにアクセスすると、オープンガバメントというのが大事で、それが透明性とか協働とかイノベーションに役立つということが、いろいろ書かれています。

ところが、これはかつてありしページでして、ある日をもってこのページはなくなってしまう。今年 1 月のことです。おわか

りのとおり大統領がかわり就任式が終わったら、数時間後にこのページが消えました。これはインターネットアーカイブからとってきたものですが、いかにアーカイブが大事かということです。やはり時の為政者によって方針は変わるといって、我々としては非常にショックな出来事ではあるのですが、かなり初期の段階からオープンデータというのがこういうことに役に立つということを、政府レベルで出していたということでも非常に貴重な資料になってしまったわけです。



〈図 9〉電子行政オープンデータ戦略

こちらは、先ほどの日本政府の電子行政オープンデータ戦略です〈図 9〉。毎年アップデートされていますが、去年の 6 月に出された最新版には、もうオープンデータ 1.0 の時代は終わりました、これからはオープンデータ 2.0 ですと書かれています。ちょっと先行し過ぎのような気もしますが、1.0 というのが行政機関が持っているデータをまず出すことをやっていこうという段階から、いかに使われるかということに目を向けて、そこを促進させていこうということが、いろいろ書き込まれています。

しかし、これをもとに各省庁なんかは、かなり意識せざるを得なくなっておりますので、こういったドキュメントなどにも目を通していただけると、動きがよくわかるのではないかと思います。

こちらが、先ほど申し上げた 2016 年末の官民データ活用推進基本法です〈図 10〉。議員

立法の基本法ということで、実はあまり内容的にもあれをしよう、これをしようということは、細かくは書かれていませんが、政府の方針として、やはりデータ活用は大事であるということを位置づけた、非常に大きな一里塚ではあります。



〈図 10〉官民データ活用推進基本法

いろいろ書いてありますが、重要なところがあるとすると、第2章です。第2章のところに、政府による官民データ活用推進基本計画の策定を、この法が施行されたらできるだけ早くやりなさいということが書いてあります。実際、5月末に政府は何をしていきますかというプランが書かれた文書が既に出てきています。

その次に、この場でも重要になると思いますが、その下に、都道府県によるデータ活用推進計画も策定しなさいとあります。これは義務であると書かれていますので、これから各都道府県レベルで、政府の計画も参考にしながらつくらなければならないということになっております。それがつくられると、当然、行政機関としてはその計画の下に取り組みが進められるということになるかと思うので、図書館等々もその影響下にあるのではないかと想像します。

さらに3行目ですが、市町村については、同じく活用推進計画の策定をするのが望ましい、努力義務ということで、絶対ではありませんが、やはり同じように計画をつくるのが望まれている。そこまでが、この法律の中

に書き込まれています。これは成立した法律ですので、いずれ何らかの形でこういった作業は各都道府県、各市町村レベルにおいても行われるものと想像されます。やはり法律の成立というのは極めてパワーのあることだと思いますので、ぜひこのあたりの動向もチェックいただければと思います。

少し日本政府の話に特化しましたが、世界的にオープンデータの流れというのは進んでいます。政府が進めているだけではなくて、それを非営利の団体がチェックするという体制も、国際的に出てきています。



〈図 11〉オープンナレッジランキング

各国政府がどれくらいやっているかのランキングをつくっている団体がありまして、これはイギリスのオープンナレッジという団体がつくっているランキングなのですが、各国のオープンデータの取り組みをはかってランキングをつけている〈図 11〉。これは去年の1位から10位を示していますが、この中に日本はありません。19位だったかな。その後さらに31位まで落ちてしまった。頑張っているのに、周りの方がもっと頑張っていて、どんどん落ちていったということです。このランキングを見せられると、行政の方がすごく悲しい顔をされる。ただ、さらに上げなければというので頑張って今年版は1けた順位になるのではないかとされています。

ただ、これも評価軸によってランキングがころころ変わるというのは当然のことですので、ここの団体では、何を公開していると点

がつくのかというのを見るとわかりやすいです。政府の予算であったり統計であったり、法律です。あるいは、Air Quality というのは空気がきれいかどうかとか、マップとか天気予報。あるいは選挙結果、水のクオリティ、政府支出などなど、そういった項目でどれぐらいの公開がなされているかを、各国のNPOなどにチェックを依頼し、その報告をまとめてこのようなランキングにしている。ルール等、日本の国情と合わない部分というのは必ずあるので、必ず何点を取るということのために頑張っても仕方がないわけです。こういった形でもグローバルに活動はチェックされているということで、このようなランキングも少しですが注目されている面があるということです。

ここまで国というレベルでのオープンデータの話をしてまいりましたが、当然、その国だけでなく都市単位、より身近な、生活に密着した形でのオープンデータを行う主体というのは、やはり自治体だったりします。

都市オープンデータ

- ・ 世界の動向
 - ・ 米国:オープンデータカタログに12都市が参加
 - ・ 英国:105/434自治体がオープンデータ協議会を設置
- ・ 国内の動向
 - ・ 2012.1 福井県鯖江市の取り組みが開始
 - ・ 2017.6 283/1765自治体がオープンデータ公開中
- ・ International Open Data Dayへの参加
 - ・ 2013 8都市
 - ・ 2014 32都市
 - ・ 2015 57都市
 - ・ 2016 67都市

〈図 12〉都市オープンデータ

アメリカやイギリスというのはかなり進んでいて、少し古い情報ですが、イギリスでは4分の1ぐらいの自治体がオープンデータをやっていますと言っていることが調査で出てきています〈図 12〉。

日本ではどうかと言いますと、最初に自治体レベルでオープンデータをやっているとはっきり言ったのは、福井県の鯖江市です。今はもはやオープンデータの日本のメッカとも

言われるぐらいになっていますが、そこが始めたのが2012年で、大体5年半ぐらいたちました。調べ方はいろいろありますが、日本には1,765の自治体があって、うち283の自治体は、みずからオープンデータをやっていますということをはっきりと明言している。ですので、1割は超えて、大体15%ぐらいの自治体はオープンデータを何らかの形で進めている。

都道府県レベルですと、ほとんど何かしらはやっているという状態になってきて、今はここまでは来ているということです。

インターナショナル・オープンデータ・デイというイベントがあって、毎年1回同時多発的にオープンデータ関係のイベントをしようということを世界的にやっています。全世界で二、三百の都市がやっていますということをやっているのですが、日本がエントリーしている都市というのもどんどん増えてきていて、最初は8都市だったのが、今は60とか70です。大体2月の終わりから3月の頭のどこかの土曜日にやっていたりします。今年はまだ2月にありましたが、来年、2018年も2月、3月ころに行われると思いますので、ぜひ注目していただきたいと思います。

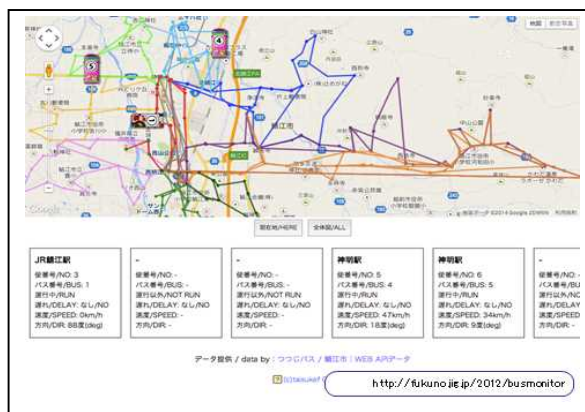


〈図 13〉日本のオープンデータ都市

これは、有志でつくられている地図です〈図 13〉。日本のどこでオープンデータをやっているかというものをつくったのですが、もはや283にもなってくると、もういたるところでオープンデータをやっているという感じにな

ってしまい、だんだん表として意味がなくなりつつあり、各都市で何の情報をオープンにしているのかということが、少し違って見られたりします。

例えば AED の場所や、公衆トイレがどこにあってそれが幾つぐらいあるのかなど、そういったわりとシンプルなデータから始まって、少しずつ情報が高度化しているというのが現状だと思います。例えば、鯖江市は市役所のページにデータを公開していますということが、2012 年の時点から出てきていて、最初は各部署が持っている特定のデータを載せていたのが、だんだんと高度化していった。



〈図 14〉バス走行情報

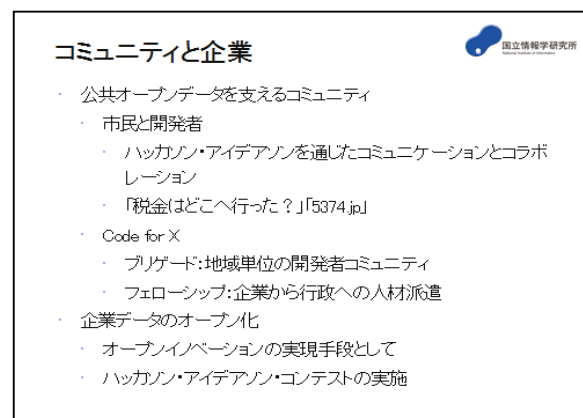
なかなかすごいと思うのは、鯖江市は市バスはないのですが、コミュニティバスにタブレット端末を設置している。そうするとタブレット端末には GPS がついていきますので、バスがどこを走っているのかという情報がリアルタイムで出る。大きな都市ですとそれぐらいのことはやっているとされていてしまうかもしれませんが、さらに何人乗ったとか何人降りたかということを経営者の人が全部押しボタンでチェックできる仕組みをつけて、今確実に 10 人乗っていますとか、3 人降りましたという情報を全部オープンにしていこう〈図 14〉。

鯖江市は人口 6 万人超の小さな市です。もちろん都市のサイズによってできること、できないことというのはあるわけですが、ある行政のワークフローの中で出すデータを超え

て、必要だったらつくってでも出すという、少しずつ変化をしているということが、非常に重要だと思います。

鯖江市が偉いのは、データを出して終わりではなくて、鯖江市にある IT 企業の方が非常にこの運動に強くコミットしていて、役所がデータを出したら、次の日には何か見せるアプリをつくるようなことをやっています。やっぱり出すだけだと、出しっぱなしで誰も使わない、何で出すんだということが必ず問われてきますが、「いや、ちゃんと使われています」とか、「これを見ている人がいます」というように説明がつくと出しやすいというのは、どこの部署であろうが同じことだと思います。そういった関係性です。実はそこが、オープンデータというのは、単に出す仕組みをしっかりとすればいいということだけではなくて、使う側の人たちとどうコミュニケーションをとっていくかということのほうがよほど大事ではないかと言われているわけです。

その流れの中で、コミュニティとどうつき合っていくかというのが行政側にも問われるようになってきている。地域にはいろいろなコミュニティがあると思いますが、特にプログラムができる人のコミュニティ、データを活用したくなる人が集まっているコミュニティが強いところは、相乗効果で非常にオープンデータが進むという傾向は、やはりあると思います〈図 15〉。



〈図 15〉コミュニティと企業

最近では、データを出したら、データを使っ

たアイデアを募るために1日集まってディスカッションをすとか、さらに踏み込んでいけば、プログラマが何人か集まってそのデータを使って1日で何か簡単なアプリをつくってしまうとか、そういったワークショップがあちこちで行われています。ハッカソンとかアイデアソンという言葉も、かなり定着してきたと思います。やはり出しっぱなしではなくて、そことの関係をつくれているということが非常に重要だと思います。

初期のある種の成功例として、「税金はどこへ行った？」というのと、この「5374.jp」というのを軽くご紹介しておきたいと思います。

「税金はどこへ行った？」というのは、各自治体が予算書、決算書を出していますので、それを見ながら、こういうツールがあって、そこにデータをきちんと入れていくと、例えば、私は年収幾らなので、多分、市民税をこれだけ払っていると思います。そうしたら、1日当たり子育てに幾ら使われているのか、道路の工事のためにいくら使われているのかというものがざっと出る。見た目を表示するためのソフトだけがオープンソースで公開されていて、データは地元のコミュニティでつくってください。そんなソフトウェアがあります。こういうのがあると、見たことはないけれど、ちょっと予算書を見てデータを入力してみようということをやられていたりします。それで、入力していくとだんだん「おや？」と思うわけです。何で我々はデータをばちばちとExcelに入れているのだろうと。そもそもここに流し込める形で行政がデータを出してくれたら、入力すらせずにいつでも最新の情報が見られるのにな、というところから対話が始まっていく。そういうのも1つの仕掛けであり、これは結構何百という自治体のコミュニティで試しにやられたという例があります。

もう一つさっきの「5374.jp」というのは何かというと、ごみなしということです。これもかなり身近な例で、燃えるごみは何曜日に

出せばいいのか、町によってルールは違うわけで、そういうものをきちんと見やすい形で表示するアプリを皆でつくろうということで、アプリが先にできて、データはそれぞれの自治体のコミュニティでつくってくださいと。これは日本製のソフトウェアで、かなり息の長い活動になっています。最初は日本語のインターフェースだけがあったのですが、住民で英語を話す方向けに英語版をつくろうとか、そういった形で展開している。そんなプロジェクトもあったりします。これも行政のごみ収集データという元データが当然あるわけですので、それをもとにつくっていく。そんなことがあったりします。

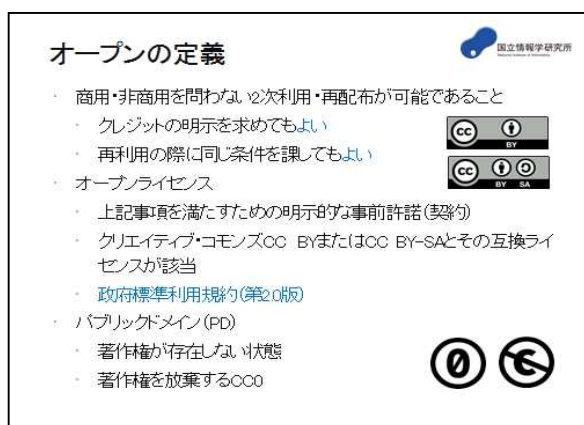
先ほど少しご紹介したインターナショナル・オープンデータ・デイは、イベントといっても10人ぐらいでやっているところもあれば、300人ぐらい集まってしまう巨大なイベントもあり、非常にまちまちです。やっていることも、シンポジウムをやっているところもあれば、ハッカソンをやっているところなど、まちまちですが、とにかく同じ日にやる。実はこういうところで人が集まり、それがまた行政からデータを出してもらう契機になっていくという、何とかそういうサイクルをつくれないうことで、皆さん頑張っているところであります。

先ほどから国の話、地方自治体の話、あるいはカウンターパートなりコミュニティの話ということで、データが今どういうふうに出ているかとしているかという話と、それが使われるためにはどうなったらいいかということについて、軽くここまでご紹介をしてきました。

この辺は、すべてオープンデータ現象というものであります。ここから先は、オープンデータが今はかなりきちんと専門用語としての定義があるという、その話に移っていきたいと思います。

最初にオープンデータというのは単にネットにぽんとデータを載せれば済むという話で

はないと申し上げました。では何を満たせばオープンになるのかというところの議論の話です。



〈図 16〉 オープンの定義

今、オープンには確固たる定義があります（図 16）。それは、国際的にもかなりの議論の末に決まったものがありまして、その内容が実際にネットで公開されています。Open Definition というページがあって、文字どおりオープンの定義なのですが、そこで書かれていることは何かということ、ある情報が、商用、非商用を問わずに二次利用、あるいは使った人が加工した結果をさらに再配布してもよいという状態をオープンと呼ぶということです。これは、考えてみるとなかなか難しい。よくありますよね、非商用だったら自由に使ってくださいとか、連絡をくださいとか、場合によってはお金を払ってもらいますよとか、そういったデータというのは結構そこらじゅうにあると思いますが、そうやって人や用途によって区別をしてはいけないというのが、1つのオープンの定義です。もちろんそれを使って、次に公開する人を差しとめすることもできないというのも、また1つのオープンの定義です。

そのときに、情報の出し手が使う人に依頼すること、課すことができるというのは唯一2つだけあります。1つはクレジットの明示を求めてもよい。クレジットというのは、要するに出し元はこの人ですとか、この人のデータを利用させていただきましたというこ

とを、使う人にちゃんと書いてということを経すことはやってもオーケーです。

もう一つは、再利用されるときに、私が出したのと同じ条件で出してください、ということをお願いすることもよい。これはちょっと複雑です。

条件を課さなかったらどうなるかというと、自分はオープンで出したのだけれど、次にそれを使って何かをする人が、商売をしてもいいということになるのです。その人が用途を区別、商用利用はだめで普通の人にはオープンにするとか、そういう条件を次に使う人が自分で課すことも、私は許すということです。ただ、もちろんそういうことをしたら、次の人はオープンデータとはいえないわけです。どう使われても自由、いいということは、仮に後でその人がぼろもうけしようが何をしようが、私はそれを許すということです。これをオープンというのは難しいということをおっしゃる方も非常にたくさんいらっしゃいますが、逆に言うとそれ以外のことはやっぱりオープンとは呼んではいけない。それは単に情報が公開されただけであって、オープンデータにはなっていないということになります。

これを実現することの難しさはさておき、ではこれをどうやって、次に実際に情報を使う人に伝えるかということも、また考えなければいけません。

大体情報というのは、つくった瞬間に著作権というのが出てくる。著作権は、情報がどう使われるかということを著作者がいかようにも決めてよいという権利のことを、一般には著作権と呼ぶと思います。その権利を持ったつくり手がいちいち「連絡してくれたら、判断して使っていいかだめかを決めます」と言っていると、問い合わせが殺到してやっていられなくなります。

これは行政の情報公開制度などもそうだと思いますが、聞かれて答えるというのは非常に大きなコストであって、そういうことをもし避けたいとするならば、あらかじめ今の、

例えばデータを利用するときにはクレジットを明記してくださいというのが書かれた契約証書を、一緒につけた状態で公開してしまえばよいということになります。その契約証書のことをライセンスと呼ぶわけです。ライセンスをはっきりさせた状態で出したい情報を公開するというのが、オープンデータの実際の進め方としていいのではないかとされています。

それで、今度はライセンスを書かなければいけないのですが、ライセンスを書くにはやはり弁護士さんに手伝ってもらったりしなければいけないので、また非常に手間がかかる。法的に間違いのない文章を書くというのはなかなか大変なことです。そこでまた1つ障壁があるわけです。

クリエイティブ・コモンズという名前を聞かれた方がいらっしゃるかもしれませんが、近年では、クリエイティブ・コモンズのライセンスを使って公開するというのが、デファクトスタンダードといえますか、これを使うとわかりやすいですという状況になっています。

そこはアメリカの非営利団体なのですが、そこがライセンスのひな形を6種類用意しています。そのひな形は、各国のボランティアコミュニティが自国の著作権法との関係をきちんと整理した上で、「このライセンス証書をそのまま使えば、グローバルにこの条件で使ってくださいということを示すことができます」というもので、それを使って公開すると、いちいちライセンスを手書きしなくてもいいというように提供しています。

クリエイティブ・コモンズの6種類のライセンスのうち先ほどのオープンの定義にきちんと合致するような内容になっているライセンスは、2つあります。

それが、CC BY と CC BY-SA という2つのライセンス証書です。BY というのが、クレジットをつけてくださいということを頼むという意味で、SA というのが、次に配るとき同

じ条件で配ってくださいということを頼むことを意味しています。CC BY-SA で配ると、次に使う人も CC BY-SA で配らなければいけないので、永久にこの条件で2次配布、3次配布という同じ条件が続いていくということを保証するという意味があります。

クリエイティブ・コモンズのライセンスの中で、この2つのどちらかで公開されたものは、オープンデータとストレートに呼べるということになると思います。ほかには非商用利用だけであったりとか、改変してはいけないとか、いろいろな条件が入ったライセンス証書があるわけですが、それを選ぶとオープンという定義からは外れるということになります。

今のところは、CC BY か CC BY-SA を使いましょと。基本的にはそうやって出せばオープンになると、わりとわかりやすくなってきたので、使われることが多いです。

ただ、政府になると、いきなりアメリカの非営利団体がつくったライセンスをそのまま使うということはなかなか難しいので、自前でライセンスをつくって、それとクリエイティブ・コモンズとの関係性を明記するといった方法をとっている国が幾つかあります。その国のうちの1つが日本であります。日本では、政府標準利用規約というのができていて、もともとは政府のウェブサイトの利用規約を定めたものですが、その中身を読むと、政府標準利用規約の第2.0版というのが2年前にできたのですが、その2.0版からは、「CC BY と同等の効力を持ちます」ということがはっきりと書かれています。ですので、これを省庁が自分のウェブサイトの利用規約として使っている、そこにデータが幾つかアップロードされているということは、そのデータはすべて政府標準利用規約に準拠している。さらに言えば、それはCC BY と同じであるということまで来ているということで、ほぼすべての政府のウェブサイトにそういう利用規約が使われているので、政府が出すデ

これはちょっといろいろな議論がある絵なので、必ずしもこの5つ星こそが究極のオー

3つ星を取ろうとすると、オープンフォーマットにする必要があります。これも意味するところの逆で、オープンフォーマットでは

ないのはどんなものかといと、これは XLS など、ある会社が自分で決めた独自のフォーマットで公開したデータというのは、その会社が突然、例えばマイクロソフトがワードとかエクセルとかをつくるのをやめたら開けなくなってしまうので、そうではなくて、国際標準か何かでデータの形式が決まってい、他のソフトでも、複数以上のソフトウェアでそのデータを見ることができる、見て再利用できるようになっているような方法で出してくださいというところを満たすと、3 つ星がもらえるということになります。

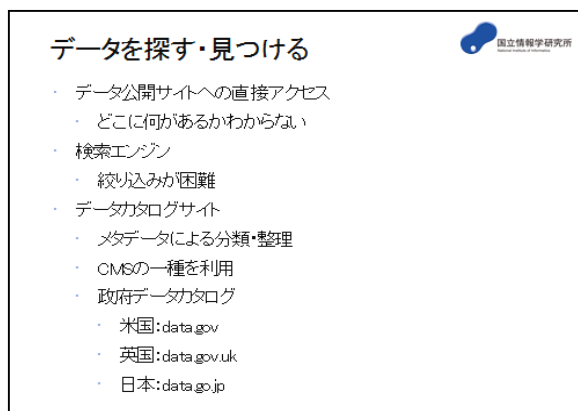
ここでは例として CSV と書いてありますが、CSV というのはカンマで区切ったテキストです。先ほどの 3.11 のときのファイルはただのテキストファイルですが、あれがまさに CSV で、この 3 つ星を満たすということになります。ほかにも XML とか幾つか標準形式というのがありますので、そういった形式で出すのがいいのではないかとっているものです。

4 つ星、5 つ星というのがある種くせ者で、ここはかなり高度なデータの表現様式を示していて、実は私はここを研究テーマにしてきた人間です。データとデータをつなげて表現するというリンクト・オープンデータとか LOD とか、そんな言葉で定義されているデータの形式があるのです。ここではほとんどその話はしませんが、CSV で出せばいいというレベルとはまったく違う情報の表現の仕方をしていくと、後々より使われやすいということを行っているのですが、4 つ星、5 つ星というのは専門的な知識が必要ですし、いらないのではないかとおっしゃる方もいます。これはもともと、ティム・バーナーズ＝リーが言っていることなので、ティム・バーナーズ＝リーがそういった情報の形式を推奨しているということではあるのですが、このところはかなり果てしない話になってしまいますので、今日は割愛します。

いずれにせよ技術としてはプログラムで処理しやすいようになっている、ライセンスと

してはそういうふうに工夫して、2 次公開、3 次公開してもいいデータというのがオープンデータであるという定義になっているということ覚えていただければと思います。

実際そういうオープンデータの定義に合致するデータが、先ほどの政府の方針とか、いろいろな方針のもとで進められてきています。数年前に比べると明らかに、手に入るデータは増えていますがこれをどうやって探すのか。実際に使おうと思ったときにどうしたらいいのかということは、まだまだ整備の途中です。先ほどオープンデータ 2.0 という言葉もありましたが、使うことをどうやって支援するかというほうに今は向かっている途中であると思っています。



〈図 18〉

データを探すのも実はなかなか一苦労で、Google でデータを探すのは結構難しいです 〈図 18〉。人口の表といっても、それは何かグラフの絵だけが出てきて数字が手に入らないといった経験があると思います。やはりそういうのは問題視されていて、どこに何があるか全然わからない。少なくとも政府の単位では、自分の政府が出している情報は、それなりの形でひとまとめに探せるサイトをまずつくることが結構大事なのではないかということで、データカタログサイトというものをつくるというのが、常套になっています。

政府レベルのデータカタログサイトというのは、アメリカにもイギリスにもあって、当然、日本にもあります 〈図 19〉。



〈図 19〉データカタログサイト

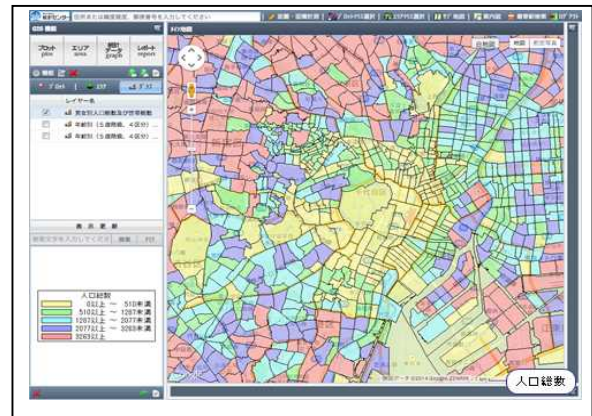
日本は Data. go. jp というサイトに行くと、例えばどの省庁がつくったものですかとか、防災に関係のあるものとか、気象に関係あるものとか、ある程度分類がしてあって、それをもって絞り込みをして、最終的に何年度の何とか調査というファイルが手に入ることができるようになっています。これも、この時点で 1 万 2,800 件になっていますが、本当はデータはもっともっと際限なくあるわけです。省庁間できちんとデータをつくり、それらを全部取りまとめて、ここに入れられるなら入れてくださいということが、ここ二、三年きちんと続けられてきていて、そこに集められたものが数万件存在するということになっています。

中身は軽く流していきませんが、例えば、国勢調査の人口データのようなものは、千代田区永田町 1 丁目、2 丁目というレベルでデータとして当然入手できるものになっています。

こういうデータを使ってアプリをつくる人というのが結構いらっしゃるって、今、都道府県別にいろいろな調査の結果を偏差値的に見せる形のサイトがあります。例えば、牛肉の消費量が一番多いのは奈良県ですが、豚肉だと北海道というように、家庭で何にお金を払っているかを総務省の家計調査結果から、見せたりしているサイトがあったりします。

あとは、総務省に統計センターというのがありまして、そこでは統計情報をかなり一点にまとめていて、便利に検索できるサイトが

できています。そのデータの使い方として、地図にいろいろな統計情報をマッピングするというサービスで、ブラウザ上で使って試せる JSTAT MAP というサービスを提供していたりします。



〈図 20〉男女別人口総数

例えば、人口を先ほどの細かい町のレベルで分類したマスに割り当てて、どこにどれくらい人が住んでいるのかなどを提供している〈図 20〉。人口総数全部を割り当てると、真ん中は皇居ですが、新宿や豊洲に人がたくさんいる。0 歳から 4 歳、いわゆる子育て世代がどの辺にいるのか、20 歳、24 歳の若者はさすがに豊洲にはいなくて、新宿や池袋などもうちょっと楽しそうなところに住んでいるとか、これは基本的な情報ですが、こういったものもいろいろ掛け合わせていくと、例えば出店で新しい店をつくるところはここがいいというようなことにつながっていく。深い分析をしようとする、手でデータをほんとうにたくさんいじっていかなければいけなくなります。単にマッピングしただけでわかることだけでは、アイデアというところまではなかなかつながっていかないとはいえ、データというのは実はたくさん出ていて、こういうところも試しながら知っていくことも重要かもしれません。

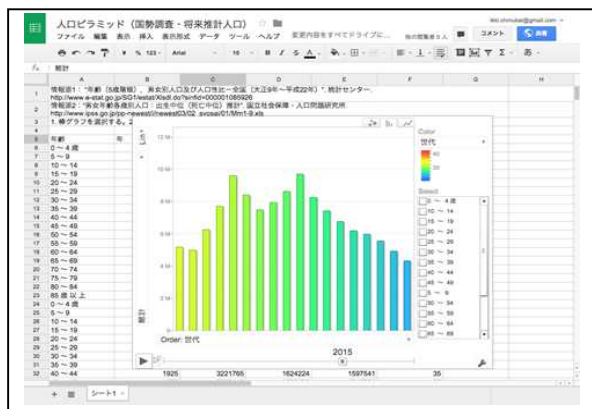
人の事例ばかり紹介するだけではなく、自分でも何かつくることができます。例えば絵はよく教科書などに出てきますが、人口ピラミッドぐらいは、当然私でもつくれる〈図 21〉。

国立情報学研究所
National Institute of Information and Communications Technology

- ということで、すべて元データはあるはずですので、自分でつくってみるということが、ある種、教育的に役に立つという瞬間もあれば、ほんとうに独自に新しい何か知識発見のために使う、つくってみる、使ってみるということが出来るかもしれない。

私がつくるためには、これまでの国勢調査の5年ごとのデータが全部手に入りますから、それを取ってきて、この先の未来は国勢調査はありませんので、そのかわりに国立社会保障・人口問題研究所がつくっている将来の推計データというのが、また数年単位で数字が出ているので、そういうデータをあちこちからダウンロードしてきて、コピペし、自分で少し整形する。

これは Google のスプレッドシートのソートを使って、あとはとにかく表示みたいなことをすると、これぐらいの人口ピラミッドがつくれて、なおかつ再生ボタンを押すと、1920 年から 2050 年ぐらいまでのピラミッドの形が徐々に変わっていくというようなことぐらいまでは、別にプログラミングしなくても、データのありかと少しの Excel のスキルは要るかもしれませんがつくっていくことができる（図 22）。



〈図 22〉人口ピラミッド

国立情報学研究所
National Institute of Information and Communications Technology

- 〈図 23〉

ただ、実際には結構いろいろなことを考えなければいけなくて、データはそれぞれ独自の基準でつくられているため、片一方のデータは 85 歳以上は全部まとめられているけれど、もう 1 個のほうは 75 歳以上がまとめられているとか、単位が全然違うなど、そういうのをきちんと理解しなければいけない。データを扱うスキルというのはやはりそれなりに必要で、そのあたりが今、データサイエンティスト人材とか言われていますが、基本的なデータのリテラシーというのはこういうものだ。出れば出るほど適当に使って変なデータを出しても仕方ありませんので、学んでいく必要があるのかもしれないということです。

国立情報学研究所
National Institute of Information and Communications Technology

- 〈図 24〉 利用条件の確認

作ったものを公開できるかどうかはもう利

用条件次第なので、もし今公開しようと思えば、それぞれの元データの利用条件が政府標準利用規約なのか、固有の規約があるのか、あるいは出してはだめとか、そういったことが書いてあると思います。そういったことはやはりきちんと見なければいけません（図 24）。

よく言われるのは、単なる数値データや事実に関する情報というのは著作権がないので、著作権とは関係ありません。人口が何人かというのはただの事実ですから、誰がつくろうが著作権はありませんので、そういうものはどんどん使っていていいとか、明示的に書いていなかったりするのが、なかなか使う側にとっては難しいところです。やっぱり著作権というのは著作物、創作です。おのれの創作の結果について権利がつくものでありますので、そうではないものときちんと見分けて、これはライセンスがなくても堂々と再発行していいという判断ができないと、なかなか次に使ってもらえる情報になっていくのは難しいと思ったりすることはあります。

元データは、カタログサイトとか、先ほどの統計センターとか、自治体もまたカタログサイトをつくっている例がありますし、市役所のサイト等で一覧が公開されている例とかもあるので、そういったものを集めながら必要なデータに当たっていくということになります。

今度はそれを活用された例はどうやって見つけるか、これが結構骨でして、今のところ、こうすると絶対に見つかるというのはありません。

データを出しているほうも、これはどう使われたのか集計したくなるわけですが、基本的にオープンに使っていいと言ってしまっているので、お礼も来ないし許可も来ないので、知らない間に使われて、その事実を出している側が知らないというのが、結構よくあります。お願いしてはいけないかということもちろんそんなことはないのですが、ライセンスにいく

ら許可を求めてはいけないと書いてあったとしても、使うときに一声かけてくれるとうれしいですということをお願いすることは自由なので、そういったことから集めたりする。

最近は、こういうオープンデータを使ったアプリケーションコンテストなど、情報の可視化を皆で競うみたいなことをやっている例もたくさんありますので、そういうところに事例は集まっていきがちであります。

ここにコンテストを幾つか上げていますが、Linked Open Data チャレンジとかアーバンデータチャレンジとか Mashup Awards とか、見ていただくと、テーマごとにいろいろなコンテストをやって、賞金を出したりしていますので、そういうところに腕に覚えのある人が集まっていたりします。そういうところを見ると、少し状況もわかってきます。やっぱりどうしても事例を完璧に迷わずに探すということが、手作業では難しいというのが、今のところの我々の悩みでもあります。

図書館でもオープンデータというのは、現実には少しずつですが始まっています。図書館は、やっぱり情報をたくさん持っている組織のうちの1つと言えると思います。もちろん本に関する情報として、書誌データだったり何冊の本を持っているか、あるいは典拠となるような名前のリストとか概念のリストみたいなものはたくさん持って、それを図書館サービスに生かしていくわけですが、こういうものがオープンデータにならないだろうかということが、一つ期待されているところでもあります。

何のために出すかという、図書館のサービスで、周りの人が勝手によくしてくれるかもしれない。スマホ版の OPAC とかつくってくれるかもしれない。今はつくってもらわなくても最初からあるかもしれませんが、便利なインターフェースをつくってくれる人も出てくるかもしれないという期待もあります。多分それ以上に、図書館のためではなくて社会に対して知識の資源をもっと使ってもらおう

じゃないかというために、オープンデータにしようということを考えていらっしゃる方もたくさんいます〈図 25〉。

書誌データのオープン化

- ・ 構造化されたデータがすでに存在している
- ・ MARCの整備(規格面・運用面)
- ・ ネットワークを通じた情報流通も実現済み
- ・ 書誌ユーティリティと業務システム・OPACの連携
- ・ 図書館コミュニティの継続的な努力の成果
- ・ 各国の取り組み
 - ・ ドイツ国立図書館・英国図書・フランス国立図書館
 - ・ 国立国会図書館
 - ・ WorldCat
 - ・ NACSIS-CAT/Cinii Books



〈図 25〉 書誌データのオープン化

図書館のデータというのは、これは完全に今までの図書館コミュニティの汗と涙の努力の上で、ものすごくきれいに構築されているわけです。書誌のデータの書き方とかが全部決まっていますし、それに基づいて入力されています。ですので、それをコンピュータ向けに加工して出すことは、ごくごく簡単に済んでしまう。それがカード目録しかなければ、それを出して使ってもらおうということはほとんど無理だったと思います。今やすべてオンライン化、データ化がされていますので、そういったデータをそのまま出すということは、少なくとも国立図書館レベルですとやって当然という感じになっていまして、ドイツ、イギリス、フランスの国立図書館はライセンスつきできちんと出しています。

国立国会図書館は、実はライセンスがまだ決まっていない状態で、データが使えるかどうか明言していません。国立国会図書館としては、別に何か文句を言うつもりはないのですが、そこはまだ明示化されていない状態です。

あとは、世界では WorldCat という書誌ユーティリティがありますので、そのデータが一部自由に使えるようになっているとか、我々が今メンテナンスしています NACSIS-CAT のデータというのも、既にクリエイティ

ブ・コモンズのライセンスをつけて公開しています〈図 26〉。

NACSIS-CATのオープン化

- ・ 制度面の検討
 - ・ 過去のデータのライセンスの主体は誰か、どのように意思決定するか
 - ・ 規約の洗い出し→該当の条項は存在せず
 - ・ 「大学図書館と国立情報学研究所との連携・協力推進会議」NIIの協議体制を確立
 - ・ データ公開の主体をNIIとする
 - ・ 書誌データのライセンスをCC BYとする
 - ・ パブリックコメント(2013.12～2014.1)
 - ・ オープンデータの公開(2014.7)



〈図 26〉 NACSIS-CAT のオープン化

ライセンスをつけるのが結構大変で、昔からあったデータに利用条件を後づけで決めようとすると、一体誰に許可を取ればいいのかという問題が出てきたりするので、どちらかという会議をいっぱい通すというのが大変だったという感じです。基本的には大学図書館全体の管理のもとで、この場合だと NII がライセンスをつけるときは何のライセンスにするかということを監督するみたいな組織的なたてつけをつくった上で、CC ライセンスを大学の書誌データについて出すということをやったりもしました。

今のは書誌とか、本そのものに関するデータの公開ですが、やはり図書館の知識というのはそれだけにとどまらず、いろいろなデータを持っております〈図 27〉。

典拠・ID・シソーラス

- ・ 典拠・IDのオープンデータ
 - ・ LC Linked Data Service
 - ・ Web NDL Authorities
 - ・ VIAF:著者名典拠データの国際連携
 - ・ 世界中の図書館で同じ作者に同じIDを与える
 - ・ 各国語版のあらゆる著作を集められる可能性
- ・ シソーラス
 - ・ dewey.info:DDCのオープンデータ化
 - ・ NDCオープンデータ化(期待！)
 - ・ MeSH Terms



〈図 27〉 典拠・ID のオープンデータ

例えば、著者名典拠なんかも、非常に管理が

されているデータではあります。それを図書館が単体で使うだけではなくて、世界中の国立図書館で典拠データを持ち寄ってどんどんつないでいこうという運動があります。

VIAF という Virtual International Authority File というサイトですが、これはある人の典拠があって、例えば村上春樹、ロシア語表記とか、そういったものというのは日本の図書館は管理していないわけですが、それもデータを集めてつなぎ合わせると、ロシアの図書館ではロシア語表記みたいなのもわかりますし、韓国語の表記も、ある個人についての表記がわかる。わかると、ロシア語版の村上春樹の本を探したい、ということがあったとしても、それをコピーしてロシアの検索サービスに入れば、とりあえず出てくるというところまではたどり着くわけです。それぞれ個別の知識だけではできないことでも、集めてつなぎ合わせると、実際に活用できたりする。そんな例もあったりします。

これは dewey の分類です。アメリカをはじめとして国際的に使っている DDC という分類があります。そのデータも公開されていますし、日本でも NDC の十進分類を公開するという方向で決まりましたので、いずれそういうデータはいろいろな形で活用できる。図書館実務のためだけではなくて、いろいろなものを実は十進分類に合わせて何か整備するということも、ほかの分野の人がやってみようと思うこともできるんだろうということです。

あとは、これは NDL です。NDL が持っている典拠のサービスです。コンピュータという概念には、上下に何があって同義語に何があるかみたいなことを表示してあります。それは例えばこういうネットワーク図を見られたり、もちろんこのデータをそのままコンピュータ処理しやすい形でもダウンロードすることができるようになっています。

例えば国会図書館典拠のビズケットという概念は、アメリカ議会図書館のクッキーに相当するとかちゃんと定義してあったり、ある

いはウィキペディアではこの言葉に相当しますとか、ウィキペディアの英語版ではさらにこうなっていますとか、そういうふうに情報と情報がどういうつながりがあるのかということが定義されているので、こういうところでどんどんデータが外の世界とつながって、新しい情報の発見とか情報の整理につながっていくことが期待されているということになります〈図 28〉。



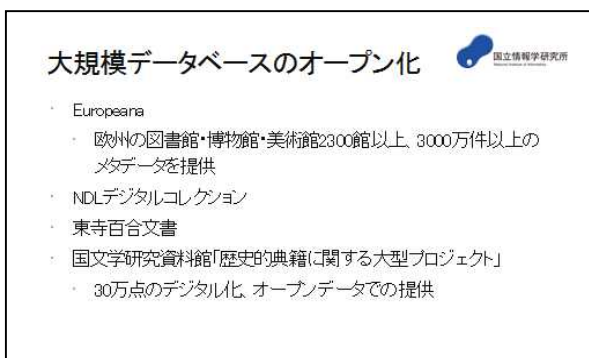
〈図 28〉 つながるデータ

そうしてできていくであろうデータの世界というのが、リンクトデータ、つながっているオープンデータという、先ほどティム・バーナーズ＝リーが言った、4 つ星、5 つ星の世界の話です。そういったものに対応したサイトというのが、これは図書館に限ってはいませんがたくさんあって、新しい情報の世界をつくっていくということをおっしゃっている。そういったティム・バーナーズ＝リーの提案に基づいて、そこによって公開するサイトというのは非常にたくさんあります。こういう中のある一群をライブラリーが占めている。この辺りとかピンクのやつなんかは、ライブラリー系のサイトだったと思います。

メタデータだけではなくて、実際に画像データをオープンライセンスで公開している例も本当にたくさんあります〈図 29〉。

一番有名なのは、Europeana と呼ばれるヨーロッパの美術館とか図書館とか文書館とかが束になって巨大なポータルサイトをつくって、例えばピカソの著作をヨーロッパ中から一度

に探することができるサイトがあります。ここで手に入る画像ファイルは、CC BY-NC じゃないとだめですよとか、そういうことが取り決められていて、その中でヨーロッパの実際のデータを持っているところは、条件を定めてユーザーに使わせるということが行われていたりします。



〈図 29〉大規模データベースのオープン化

国立国会図書館もデジタルコレクションというのは持っていて、近代のものをスキャンして公開しています。そこもまだライセンスについてはきちんとしていない部分があります。そのかわり著作権が明らかに切れたものは著作権切れということをはっきりと明記するようになっていきますので、その部分に関しては明らかにオープンに使えるということがわかっているということになります。

あと、数年前に公開されて非常に話題になったのが、この京都府立京都学・歴彩館の東寺百合文書というのがあります。これがすべてデジタル化されて公開されていますが、自治体系のデータベースでいち早くすべてのデータに CC BY のライセンスをつけたということがあり、これをもとにライブラリーオブザイヤーの大賞をとったり、非常に注目された事例であります。

あとは、立川市にある古典を研究するための研究所である国文学研究資料館では、持っている 30 万点の古典籍を全部デジタル化し、CC ライセンスをつけて、今、順次、公開中です。今、3 万点ぐらいで、ここから数年かけて徹底的にスキャンをしていく予定ですが、

古典籍に関してはここが圧倒的な物量を持っていますので、それがこれからことごとくオープンデータ化されていくということで、一つの期待のプロジェクトとなっています。

図書館ですと、これは今年の 3 月でしたが、大阪市立の図書館デジタルアーカイブが、きちんとオープンデータのルールにのっとって公開したということで、今年のトピックとしては、かなり注目されています。多分、これから数限りなく生まれていくと思われ、公開方法としても、CC ライセンスをつけようということはかなり理解が進んできています。逆にライセンスを自分たちで考えることのほうがはるかに大変ですので、いずれにせよ実際にオープンに使ってほしければこうやるというのが、基本的にはできてきているのではないかと思います。

こういう事例は多分、来年になればまた増えていくと我々としては期待しておりますし、やはり使われていく。アーカイブ化を考えていくと、ライセンスのある情報の出し方というのは非常に重要になると思っております。

一方、大きいところはそうやってできますが、小さいところは何かできることがあるんだろうかという議論がやっぱりあります。



〈図 30〉ある対話

知り合いのフェイスブックでのお話ですが、山中湖にある図書館の館長の丸山さんが、大きいところはいいいけれど小さいところは何かやることあるんだろうか、ということをぼそっとつぶやいたところ、そんなに肩ひじ張

らなくていいんじゃないのと〈図 30〉。例えば、図書館のサイトで父の日にあげたい本のリストとかを公開しているじゃないですか。これをただアップするだけではなく、これにライセンスをつけたらオープンデータですよと言ったら、「ああ、そうなんだ」とか、「こんなんでいいのか」というので、逆にびっくりしている。じゃあちょっとやってみましょうという感じで、次の日オープンデータにされたことがあったんです。

実際、図書館でいろいろなことをやっていきますよね。いろいろな本を特集したりということはやっているわけで、しかもそれはネットに公開されている。こういうところはもはや事実なので、このレベルにおいては基本的には著作権は問われないわけですが、それをまとめて、これをおすすめしますというリストそのものには、誰の権利の処理も必要ないし、逆にこちらから出すにあたってオープンライセンスをつけておくということをすれば、それも一つのオープンデータである。

ということで、進め方というのは多分、かなりいろいろな方策があると思います。もちろん非常に大きなデジタルアーカイブを持っているところもあると思いますが、いずれにせよ、きっと図書館の用途の中には、かけがえのない情報、かわりのきかないここにしかない情報というのがたくさんあって、実はそれを出していく、あるいはそれを市民の方々に見せていくということが、次のアクションを生み出していくということがたくさんあるんだろうと思います。

レファレンスも、レファレンス協同データベースに入っていますが、あれもまだオープンデータとは言ってはいません。こういうところでレファレンスで出てきた本ですというリストにしてもオープンデータになるポテンシャルはあるわけです。一時的につくって、リアルな場所ではもう次の特集に行ってしまうのだけれど、残すという意味でもリスト化してあればそれを出してみたらどうです

かなど、結構やることの手の打ち方というのは、たくさんあつたりします〈図 31〉。

図書館のオープンデータ

- ・ 業務の中にある「かけがえのない情報」をオープンに
 - ・ 地域・郷土資料の書誌情報
 - ・ OPACから引き出し、使えるように
- ・ レファレンス事例
 - ・ レファレンス協同データベース経由
- ・ イベント・展示
 - ・ 本のリスト・写真…
- ・ 一時的に作られる情報をアーカイブに、資産に
- ・ 図書館自らが新たな情報をつくりだす
 - ・ ししまろはん・残念な日本地図
 - ・ Wikipediaタウン

〈図 31〉 図書館のオープンデータ

最後のパートですが、持っているはずの情報を出すということからさらに一步踏み出していくと、図書館みずからが新しい情報をつくり出すということをやったって当然構わないわけです。これも一つ注目された活動で、京都府立図書館の方が自主勉強グループでオープンデータを勉強していたのですが、実際ちょっと自分たちでもつくってみようとなりました。本を片っ端から読んで京都が出てくる本の、地名と緯度経度とその本のページ数だけをひたすら記録し、それをオープンデータとして公開するというをやって、結構注目を集めました。

ひたすら読んでいって、数としては何百件という感じなのですが、そうやってデータが出てしまえば、また誰かが、じゃあちょっと地図にマッピングしてみましようということはやったって構わないわけで、交差点ごとに事件が起こっているみたいなことがよくわかるというわけです。

そういうことを試しにやってみたり、さらに本の中で出てくるレストランの描写だけを何パターンか取り出して、この本で何という店が取り上げられて、看板メニューはこれというようなデータを、また本をひたすら読んでやっている。もちろん規模では全然大きくはならないですが、これとて楽しくやっていることは、また誰かが使って広がっていくか

もしれない。そういう例はやっぱりあるのではないかと思います。

これを最後にしたいと思いますが、図書館がオープンデータにかかわる例として、非常にいい例だと思うもので、ウィキペディアタウンというのがあります〈図 32〉。



〈図 32〉ウィキペディアタウン

既に御存じの方もいらっしゃるかもしれませんが、よく町歩きのイベントを地域でやられていると思いますが、それはやってみて終わりではなくて、その結果をきちんとウィキペディアに書きましょうと。そうすると残って、なおかつ新しく誰かが検索すると、ウィキペディアは上のほうに来るので、やった結果が活用されていく。ここまですると、参加者のモチベーションも高いということで、今非常にやっています。もともと横浜でやったのが日本では初めてだったのですが、今全国で何十、何百とワークショップが起こっていて、ウィキペディアの中にウィキペディアタウンというページがあって、そこで次にどこでウィキペディアタウンをやりますみたいなのがどんどん出てきているのですが、わかっているだけで、来月で四、五回どこかでやっている。

何が図書館に関係あるかというと、町歩きをしたり何か調べることというのは、普通の町歩きでもできるのですが、最後にウィキペディアに書くときに、適当に書くとすぐに消される。ウィキペディアのなかなか怖いところで、やっぱり信頼性の高い情報をいかに残

していくかというところで、裏がとれていない情報を書くと、それは何の証拠もありませんと言って管理者が消してしまうということが結構よくあります。

逆に言うと、裏さえちゃんととれていれば残るわけです。町歩きした後図書館に皆で集まって、そこで資料をいろいろ調べながら、見てきたものの裏をとった上で、参考文献にしてウィキペディアに書くというところまでやるというのが、ウィキペディアタウンのみそです。

そうすると、まさに司書さんの出番です。多くの場合は、「今日はこのページを書きましょう」と決めた上で始めますので、司書さんが事前に自分の図書館から関係していそうな本を棚いっぱいにとってくる。それで一たん外に出て、帰ってきてそれを見ながら何か裏がとれないかということを調べたりする。もちろん情報が足りないときはまた館内資料を探しに行くということもやりながらということなので、ウィキペディアタウンというのは、ぱっと見、図書館とあまり関係なさそうですが、大抵の場合は、司書さんはほぼ必須という感じになってきます。あとは、ウィキペディアの執筆者コミュニティというのがありますので、そういう人から、ウィキペディアではどういう書き方が望まれているかみたいな作法の説明があるので、ウィキペディアの編集者と司書さんと町の専門家、あるいは郷土史家的な人が中心になって、市民の人たちと一緒に学習する。かなり大ヒットのワークショップ形式で、何でこれがオープンデータに関係するかというと、ウィキペディアで書かれた情報は、絶対に CC BY-SA で公開されると決めてあるからです。だから、ウィキペディアの情報は 100%オープンデータなので、結局こうやってやるのが、オープンデータを生み出していくという一つのやり方です。非常に注目されているワークショップですので、これはお近くでやられている可能性が高いですし、やりたいという方がいらし

やったら、こういったコミュニティに声をかければウィキペディアの編集者が出てきてくれることもありますので、ぜひ検討していただければと思います。

これでおしまいにしていきたいと思います。オープンデータをやりたいと思ってからも、実際に行政の組織の現場では、なかなか大変なことがあるのは間違いない。世間で調査されたデータを公開したくない理由というのをスライド5ページにわたってのせていますが、どこでもそういうことはあるということです。

データを公開しない理由(1/5)

- ・ 無関心
 - ・ 誰も気にしていない
 - ・ コミュニティからの支援が得られない(市民ハッカーがいらない)
- ・ 混乱
 - ・ 明確な事例がない
 - ・ データを可のために使うのかわからない
 - ・ どんなデータが必要とされるのかわからない
 - ・ データを公開する前に、どんな便益があるか明確にしたい
 - ・ 今のシステムで十分だ
 - ・ 人々を混乱させるだけだ

〈図 33〉データを公開しない理由(1/5)

データを公開しない理由(2/5)

- ・ 困難
 - ・ とにかく複雑で難しい
 - ・ 大量のバックログを抱えている
 - ・ ついていけない、時間が足りない、変化が早すぎる
 - ・ データの量が膨大である
 - ・ 自分の時間の大部分が取られてしまう
- ・ 費用
 - ・ 予算がない
- ・ スタッフの問題
 - ・ それは私の仕事ではない
 - ・ 私の上司や政治家はデータ公開を望んでいない
 - ・ やり方をわかっていない人が多い

〈図 34〉データを公開しない理由(2/5)

データを公開しない理由(3/5)

- ・ 合法性
 - ・ データや記録を公開する法的な権限を持っていない
 - ・ このデータを公開すれば、市民は私たちを責める
 - ・ 私たちはデータに対する権利を持っていない
 - ・ 私たちがデータの所有者なのかどうなのかわからない
 - ・ データを販売するべきである。もし売らなければならぬとすればよいのか？すでに売っている
 - ・ 機密データではない、が、公開することでよいIPRになるとは思えない
 - ・ 私たちの弁護士は適切なライセンスを付与するべきだと言っている
 - ・ このデータを集めるのに、膨大な時間・人・資源を投入した
 - ・ なぜデータを必要としているのかを知らないで、直接窓口に来てほしい
 - ・ 誰かがバックアップして勝手に売ってしまうのではないのか
 - ・ データが盗用されて利用して、何者だったのかをコントロールしたい
 - ・ 連邦政府や他の政府機関がすでにデータを持っているので、そちらから入手してほしい

〈図 35〉データを公開しない理由(3/5)

データを公開しない理由(4/5)

- ・ 正確さ
 - ・ データを一括して公開すると、だれかがそれを変更してしまう
 - ・ データを公開したりすると、私たちがハックされる
 - ・ データを公開すると、やり方によっては人々が誤解する危険がある
 - ・ 市民が混乱する、データの品質がそもそも高くない
- ・ プライバシー
 - ・ プライバシー保護の懸念があり、データを公開できない
 - ・ もう既に公開している
 - ・ 窓口やプロバイダーを通じてすでに公開している
 - ・ 情報公開制度があるではないか
 - ・ すでに十分にデータは公開している
 - ・ APIを提供しているではないか
 - ・ データポータルが必要なら、同じようなものはすでにある

〈図 36〉データを公開しない理由(4/5)

データを公開しない理由(5/5)

- ・ その他
 - ・ データが優秀な技術者集団や高等教育を受けたコミュニティに利用されると、デジタルデバイド問題がさらに深刻化する
 - ・ 今、それは優先事項ではない
 - ・ それについてはベンディング状態だ、似たようなプロジェクトもすでに始まっている

出典: Reasons to Not Release Data, Part 1-Part 10, Sunlight Foundation
(日本語訳:「公的機関がデータを公開しない理由」Open Knowledge Foundation Japan)

〈図 37〉データを公開しない理由(5/5)

ですけれども、いきなり大々的に始めると、すごく大変なので、やはり小さく始めてどんどんコミュニティにぶつけて評価を得ていくというやり方をするのがいいのかなと思います。そうすることによって、図書館外の世界とつながって何らかのデータを出していくとか、先ほどのウィキペディアタウンみたいにデータをつくることに図書館が協力していくとか、そういったものの可能性が非常に大きいかと思ってきているところです。ここまでオープンデータは何かということから、あるものを活用する、図書館が持っているものを出す、それに加えて、コミュニティを支えて、新しいデータをつくり出すというところで、図書館のかかわりについてお話をさせていただきました。どうもありがとうございました。(拍手)

— 了 —