

展望	1	公開月例研究会講演記録〈第 286 回〉(2015.11.24)	25
公開月例研究会講演記録〈第 284 回〉(2015.6.18)	2	公開月例研究会講演記録〈第 287 回〉(2015.12.3)	39
公開月例研究会講演記録〈第 285 回〉(2015.7.9)	14	産研だより	60

公開月例研究会講演記録〈第 284 回 (2015. 6. 18)〉——

統一テーマ：「ビッグデータ時代における産業・企業の課題」 「Linked Open Data (LOD) の動向と今後の展望」

大阪大学産業科学研究所
古 崎 晃 司

ご紹介ありがとうございました。大阪大学の古崎と申します。

大概先生からご紹介いただきましたように、私は LOD に関する研究もやっているのですが、専門分野はオントロジー工学です。オントロジー工学の基礎理論、それに基づくオントロジー構築・利用環境の開発、特にライフワークとして「法造」というオントロジー構築のためのツールを大学生のころから 20 年以上つくり続けています。もともとコンピュータで意味を扱うという流れから、Linked Open Data のほうにそれを応用しようとして研究している人間だと思ってください。

本日の講演内容は、Linked Open Data (LOD) の基礎と、その動向、そして今後の展望ですが、まず「リンクトオープンデータ」という言葉を聞いたことがある方、どれくらいおられますか……。

ありがとうございます。「ビッグデータ」という言葉は皆さんよくご存じだと思いますが、「LOD」はまだまだのようですから、きょうは LOD の学術的・技術的背景について概要的なお話をします。本日のキーワードとして、Linked Data (LD)、Linked Open Data (LOD)、オープンデータ (Open Data)、オントロジー (Ontology)、この四つは最低限覚えていただいて、技術的な細かいことをもう少し知りたい方は、セマンティック Web (Semantic Web)、RDF、参照解決可能、SPARQL、こういう言葉も併せて覚えていただければと思います。

皆さんはほぼ毎日、Web (インターネット) を

使っていると思いますが、Web を構成する基本的な技術は URL と HTML です。よく見る `http://` 何々というのが URL で、ホームページのアドレスです。ホームページを見ると絵とか写真とかきれいになっていますが、あれはワードかなにかで書いて張ってあるのではなく、HTML という言葉で書かれています。ホームページを見て、青いアンダーバーを押すとページが変わりますね、あれをハイパーリンクと言います。

細かい技術的なことは知らなくても、パソコンとかスマホで `http://` 何々と書けばホームページが表示され、青い線のところをクリックすると違うページに行く。それがインターネット、Web の基本的な技術で、創始者はティム・バーナーズ・リーさんです。

皆さんは日々、パソコンやスマートフォンを使っているので、大した技術ではないと思っているかもしれませんが、実はこれは 25 年前にはなかった革新的な技術です。当時はパソコンでインターネットを使う人口も非常に少なかった。それがいま、`http://` 何々と書くだけで、その URL のページが出てくる。世界じゅうのインターネットに接続されて、地球の裏側のブラジルのコンピュータの情報でも自由に閲覧できる。しかもハイパーリンクによって他の Web ページとつながっていて、関連するリンクを延々とクリックし続けるようなことができていくわけです。きょうは時間の関係で技術的な細かい説明はできませんが、そのようなものがいまのインターネット社会

の根幹にあることだけ知っていただきたいと思います。

いま世界じゅうの人が普段使っている Web から、セマンティック Web へという流れになっています。いま皆さんが使っている Web は、ニュースなり論文なり解説記事なり、人が読んで理解することが前提でつくられている。それをコンピュータが理解可能な Web へということです。

これだけホームページにたくさんデータが載るようになると、もっと賢く使いたい。たとえば「藤沢」は人の名前なのか地名なのか、人が見れば分かりますが、それをコンピュータが意味を理解するようにしたい。人が読めば分かることを、コンピュータで意味処理させよう。そのためにはどうしたらいいかというのが、次世代 Web と言われていたセマンティック Web でした。これはいま、ほぼ過去形になっています。

たとえばレポートを出すときに、自分の興味があるホームページを選んで読んでいきますと膨大な資料が集まる。その際、コンピュータがホームページに書いてある内容を理解して、これは大事だとかレポートに書くとかまずいとか判断して、必要な情報だけ出してくれるとありがたい。単なる文字情報ではなくて、コンピュータによる意味処理に用いる「メタデータ」をつけていくことで、従来の人を読むための Web から、コンピュータが理解可能な Web へ、セマンティック Web へという研究を、やはり先ほどのティム・バーナーズ・リー氏が十数年前からやっています。

コンピュータに理解させようと思うと、意味が分かるだけの簡単なタグをつけるだけでも大変で、10年ぐらい前に考えられていたのは SNS とか Web2.0 です。Web2.0 と言っても分からないかもしれませんが、いまツイッターとかいろいろなものにデータが集まっているので、検索するとビッグデータが出てくる、ああいうもの先驱けです。

とりあえずみんなでどんどん書いていくとデータがたまるから、それを共同構築しようということが10年ぐらい前から始まった。それでは大変だからやり方を変えようということで2007年ごろに Wikipedia が出てきて、その流れで出てきたのが Linked Data です。セマンティック Web は2000年代当初から始まっていますが、みんなが真面目

に書くのは大変なので、もう少し楽な方法がないかということで DBpedia や Linked Data が出てきて、Linked Data という言葉が頻繁に使われるようになったのは2007年ぐらいからだ、ということだけ理解してください。

Linked Data で最初に大きなものとして出てきたのは DBpedia です。皆さんよく Wikipedia を使おうと思います。あれはかなりのデータが集まってきているので、そのままデータとしてコンピュータで処理しやすいかたちで公開すると Wikipedia のデータベースができると考えたグループがいて、そこで公開されたのが DBpedia という基盤サイトの Linked Data のはしりです。

同じような考え方から、新しくホームページを書くときにデータをつくらっていくのは大変なので、いまあるデータを公開してお互いにつないでいけばよいのではないかと提唱されたのが、やはり Web 創始者のティム・バーナーズ・リーさんで、それが Linked Open Data の始まりになります。

「データをつなぎましょう。データをつなぐと新しい価値が現れるのではないですか」というのがティム・バーナーズ・リーの基本的な考え方で、この考え方を世界的に訴えるために彼が最初に言ったのが「Raw Data Now!」です。

Raw Data はコンピュータの用語で、生データ、手元にあるデータそのままという意味です。それを Now, いますぐ Web に置きましょう。皆さん日々、エクセルでたくさんデータを集めたり、持ったり張ったりすると思いますが、それを Web に置いてみんなが使えるようにすればすごいことが起こるのではないかと、ということをして2009年2月に訴えたというのは有名な話です。そして翌年2010年2月、そうすると何が起こるか、LODの活用事例として幾つか紹介しています。

(資料1) その一つの事例として有名なものですが、白と黒の家は、白人と黒人、どちらが住んでいるかというデータをプロットしたものです。青い線は水道管がどれだけ整備されているかというデータです。それぞればらばらのデータとして存在していたものを一つの地図に表してみると、白い家の周り黒い家の周りど水道管の整備状況が明らかに偏っていることがクリアになった。これって行政としてどうなのというので、裁判の材料に使われたわけです。

家のデータだけであれば、黒人と白人と、住んでいるところが分かっているんだぐらいのことしか分からない。そこにほかのデータを重ね合わせると、また違った事実が見えてくる。データを組み合わせ、つなげることによって、新しい価値が出てくる例として、よく取り上げられています。

普段使っているいままでのインターネットは、文書を Web に公開して誰でも見れるようにして、ハイパーリンクでクリックすればつながるので、すごい利用がされている。その次の段階として目指すのは、文書ではなくデータでそれをやろう。データを公開して共有し、お互いにつながることによって新しい価値を生み出そう。それが Linked Data という考え方です。

詳しくは『Linked Data – Web をグローバルなデータ空間にする仕組み (Tom Heath, Christian Bizer 武田英明監訳 丸善 2013)』というよい本が出ていますので、ぜひ読んでいただければと思いますが、サブタイトルの「Web をグローバルなデータ空間にする仕組み」がまさに Linked Data が目指すことを表している。「グローバルなデータ空間」というのはコンピュータ屋にとっても非常にすてきな言葉で、いま全世界でインターネットで公開されて全データが一つのデータベースとなって自由に使えるようになったら、すごいことができそうだねという話です。

(資料2) そんな理想論としては分かるが、実際どうなのと思われるかもしれませんが、実際に広がっています。これは2007年5月、ティム・バーナーズ・リーさんやそのコミュニティが「Linked Data やろうぜ。いまあるデータをとりあえずつないでみよう」というので始めたもので、1つの丸がそれぞれ個別に公開されたデータベースを表していて、それぞれこんなふうにつながっていることを表している LOD クラウドと呼ばれるものですが、非常に少ないですね。

それが2007年10月にはこれだけに増えて、1年後はこれだけ、さらに2009年、2010年、2011年、最終的に2014年にはこんな感じで明らかに増えていることが分かります。少なくとも1014のデータベースがいろんなかたちでつながって、こんなかたちになっていて、さらに増えているという状況です。

カテゴリー、分野ごとに色分けしているんです

が、政府関係のデータが183、出版関係が96、ライフサイエンスの分野が83とか、いろいろな分野の情報がそれぞれたくさん公開されて、真ん中にあるのが DBpedia です。DBpedia で相当の部分がカバーされているので、ここを中心につないでいる。

決して理想論ではなくて着実にデータがつながってきているし、他のオープンデータとつながることで新たな価値が生まれようとし、一部生まれて使えるような状態になってきているというのが現状です。LODってどうなのといったとき、とりあえずこの図を思い出していただいて、オープンデータが相互につながっていることはすごいことだということを頭の片隅に置いていただければ、きょうの講演の半分は OK という感じです。(資料3) それではデータがつながると何がよいのか。Linked Data とか Linked Open Data そのものではないんですが、データがつながることによるメリットとしてよく語られるのが、知識がつながったグラフ、Knowledge Graph です。Knowledge Graph はフェイスブックやいろいろなところも言っていますが、有名なところでは2012年に Google がつくって実際使えるようにしてサービス開始したのが最初です。

実際どんなものか、Google のホームページで Knowledge Graph を説明しているところをお見せします。

いま Linked Open Data でやっていることはこれと同じで、あらゆるデータをこんなかたちでつないで意味を辿れるようにしてやると、何か本当に知りたいことをピンポイントで探し出すことができるようになるし、その一部は Google の検索でも実現されている。

(資料4) 普通の人からすれば、ある日、Google を検索するとこんなのが出るようになったぐらいにしか考えないでしょうが、「村上春樹」でクリックすれば、ここは大したことはないんですが、このへんが全部著書と分かっているというのは、あらゆるデータが整理されていて、クリックすれば、「ノルウェーの森」だったら、こういう映画があって、主演は誰だとか、全部辿れるようになっていて、それはバックグラウンドに膨大なデータが整理されているんだということをぜひ知っていただきたいと思います。

ここまでをまとめますと、従来の Web は一般化されて、最近の中学・高校生はスマホで Web 検索しながらレポートを書く、分からないことがあれば教科書よりスマホで見るというぐらいに普及しています。これからは文書の Web から人がクリックして迎えるのではなくて、機械が意味が分かったうえで、本当に欲しい情報を思いのままに引き出せるような Web を目指している。それがセマンティック Web であり、Web2.0 から、いま Linked Data に進んできている。

Linked Data のすごいところは、データをつなげることで、どんな新たな価値が生まれるかまだ分からない。Knowledge Graph で辿っていく中から思いがけない発見があるかもしれないし、データとデータを重ね合わせといままで見えていなかったものが見えてくることも今後期待されている。

そのための大前提は、Linked Data と言われるように、データとデータをつなぐことがこの技術の根幹なんだ。ホームページとホームページをハイパーリンクでつないでいるだけでも、いろいろなことができる。それがデータとデータがもっと細かいレベルで正確につながったときに、そこから新しいいろいろなことができるだろうということ想像していただければなというのが、最初のパートでお伝えしたいところです。

次に Linked Data 技術の概要についてお話しします。

(資料5) Linked Data の四つの条件というのがあります。ここで URI という言葉が出てきますが、先ほどの URL とほぼ一緒と考えてください。そのうちの1と4を見ていただきますと、「全てのモノやコトに URI をつけましょう」、「多くのモノやコトを発見できるように、外部へのリンクを含めよう」と書いてあります。ホームページを公開しようと思えば、アドレスを決めて、自分のホームページはこれですと皆さんに伝える。それと同じように、自分のデータにアドレスをつけて、自分のデータと他のデータをつなぎましょうということです。

(資料6) そのときにつなぐ仕組みとして RDF の例を示しています。「大阪市」をクリックすると、大阪市は大阪府にあって、面積は223km²で、人口は2687万人、隣には豊中市、吹田市などがある。市の木はサクラで、サクラはバラ科で、英語

名は cherry blossom だとか、そういうのがつながっているのが見えるような仕組みになっています。この1個1個の丸は、皆さんがホームページをクリックするのと同じように、ホームページに相当するデータ版だと思っていただければいいのですが、それがそれぞれ線でつながっている。これが Linked Data の基本的な仕組みです。

丸で書いているところがリソースと呼ばれたり、関係のところがプロパティと呼ばれたり、文字列のところはリテラルとか専門用語では言われていますが、URI を使って、大阪市はこうだよ、大阪市と関係があるのはこういうものだというのを書いていく。そのときに、大阪市が主語で、隣接する自治体は豊中市だとか、主語・述語・目的語という英語の基本構文と同じように書いていくわけで、これを使えば大体何でも書けます。

ですから、Linked Data はデータをつなげるのもので、技術的には RDF というこんなかたちでデータのつながりが書かれているものだとご理解いただければ、一般の方のイメージとしては十分だと思います。

見た目は単純ですが、データをつなげるという観点からは非常によい仕組みになっていて、主語・述語・目的語というかたちで書けば大体のことは柔軟に表せる。三つ組(トリプル)により、さまざまな構造のデータのつながりを柔軟に表現できることがまず一つです。

もう一つは、データを URI で表すことで、外部のデータとつながることができる。世の中に存在する全てのデータをグローバルに一意の ID で表すというのが URI のすごいところで、Web サイトの URL と同じ仕組みです。単なる数字を ID とすると、異なるデータベースが同じ ID を使っている可能性がある。それを http:// 何々とすれば、間違いなくそのデータを同定できるのも URI のすごいところです。

これが担保されているから、外部のデータと正しくつなぐことができる。先ほど見ていただいた Linked Data の四条件のうち1と2はその意味で非常に大事で、全てのコトやモノに URI をつけることによってグローバルに一意の ID が担保されている。それでつないでいけば、LOD クラウドや Knowledge Graph のように、データがどんどんつながって大きなものになっていく。このこ

とはぜひ覚えてください。

それではデータをつないだ場合、どういうふう
に公開しますかという話です。コンピュータが得
意な人は参照解決可能な http URIs を用いた公開
をしていただくと、URI でデータにアクセスが可
能になります。

実際どんなものか、例を見ていただきます。た
とえば DBpedia のページだと、日本大学の URI
はこうなっています。日本大学に関するデータ
がこんな感じで取れます。これだけと言えばこれ
だけですが、それをブラウザに入れるだけでちゃ
んとデータが出る、クリックするとつながって
データが出る。そういうあたりまえのことがちゃ
んとできるようなかたちで公開するというのが参
照解決可能な公開と呼ばれるやり方です。

いま世の中のいろんな Web ページを分析して
ビッグデータとして扱っています。たとえば
Google も、リンクされている全ホームページの
情報を毎日コンピュータで集めてきて、リンクの
され方から、どこのページがたくさんリンクされ
ているか、どの順番に出すかとか解決している。
それと同じようなことをするにはデータもそんな
つながりでやらないとだめですよというような仕
組みで、国会図書館のデータもそのようなかたち
で公開されています。

Linked Data ブラウザを使うと、データをうまく
辿っていけるような仕組みになっていますが、
時間がないので省略させていただきます。

それが一つのやり方ですが、もう一つ、コン
ピュータのプログラミングをする人向けの標準的
な検索用言語として SPARQL というのがありま
す。(資料7) たとえばこういう書き方をすると、
東京都につながっているデータが全部取れるよ
うになっています。

実際どれだけのデータが検索できるようになっ
ているか知りたい方は、関西で開催している
LOD ハッカソン関西の Web サイトに行ってい
ただくと、SPARQL 検索が可能な日本語 LOD 一
覧とか、SPARQL の使い方の解説やサンプルなど
もあるので、学生さんで実際使ってみたい人は
ぜひ見てください。

以上をまとめますと、Linked Data 技術の四つ
の基本原則といっても、一番大事なものは、全
てのデータに URI をつけること。その URI に入力

すると、ちゃんとデータにアクセスできるよ
うな仕組みで公開すること。できれば検索の仕
組みをちゃんと提供しておいて、自分のデータ
だけでなく、他のデータとつなぐことで新しい
価値が出てきますよということを満たすこと
も大切になってきます。これらの原則はイン
ターネット標準化団体による標準技術が厳
しく決められていますので、それをうまく
使っていくことが大事で、さらに技術的に
詳しいことを知りたい人はそれなりの勉
強をしてください。

Linked Open Data がどういうものか、技
術的にどういうバックグラウンドがあるのか
というお話をしましたので、次に、国内を
中心に、Linked Open Data の最近の動
向をお話します。

Linked Open Data は最初海外で、特に
生命科学とか学術情報の分野で先行して
行なわれていました。創薬や病気の解明を
するときに、全ての病気を一つの研究施設
で解決することはできないので、実験した
データをお互い公開して、データベースと
つないで実験データを共有することが昔
から行なわれていました。それが2007年
ごろから、オープンガバメントだ、オープ
ンデータをやるぞとアメリカやイギリスが
言い出して、そこから LOD を活用しよう
という動きが盛り上がってきたというのが
海外での動向です。

日本もそれにやや後れてスタートした
のですが、海外と同じように生命科学や学
術分野での取り組みが先行しています。私
も関係している2011年の「LOD チャレ
ンジ」以来、オープンデータに関する多
くのコンテストが開催されるようになります。
2013年にはアーバンデータチャレンジ、
さらに経産省や総務省など国もコンテ
ストをやるようになる。われわれがコ
ンテストをやった2011年当時、幾つか
の自治体で先行して取り組みを始めて
いましたけれども、中央政府はほとんど
やっていなかった。それが2013年の
G8サミット首脳宣言で「オープンデー
タ憲章」に合意したことを機に、2014
年に日本政府のオープンデータカタ
ログサイト本格運用が始まる。2013
年から2014年にかけて多くの自治
体でオープンデータの取り組みが開始
される、というのが日本のオープンデー
タの大まかな流れです。

Linked Open Data と Linked Data とは
Open がついているかどうかの違いです
が、技術的には

Linked Data から始まって、オープンデータに使おうということで、大きく広がってきた、というのがざっくりとした言い方になります。

オープンデータの細かい定義は省略しますが、商用でも商用でなくても誰でも自由に使えるデータです。Linked Data は相互にリンクされたデータで、RDF を用いて表されます。Linked Open Data は Linked Data を使ったオープンデータです。

オープンデータの定義は誰でも自由に使える形で公開されているデータですが、二つの重要な観点があります。一つはライセンスです。商用、非商用にかかわらず、自由な使われ方をするのがオープンデータだと言うと、公開したデータを誰かに悪用されないかと言う人がいます。しかし、考えてみると、悪用するような人はライセンスを守らない。ということは、ライセンスで縛ることは、真面目にライセンスを気にして使おうとする人にはマイナスに働くけれども、悪用しようという人にはマイナスに働かない。したがって、ライセンスをオープンにすることは、真面目な人が安心して使ううえで大事だという考え方です。

二つ目は機械可能な形式ということです。プログラムを書いたことのある人は分かると思いますが、使いにくいデータは使いたくない。簡単に使える、プログラムで処理しやすいフォーマットが望まれるわけです。

有名な話で、東日本大震災の後、東電から細かい計画停電の予定が送られてきて、サーバー管理者は大変だったと思います。しかもそれが PDF になっていて、電力需給がどの程度なのかよく分からない。東電のホームページならグラフで見れるけれども、グラフでは何%か分からない。そのときに、画像で張られたグラフを解析して、これは何%だということをデータとして公開した Web のプログラマーがいる。それによってスマホアプリでも電力需給が分かるようになったという話があります。

この例で分かるように、データは使える形で公開してもらわないとアプリケーションをつくれな。その意味でフォーマットは大事です。PDF で公開されても検索もできない。ものによっては、エクセルでデータを整理して、スキャナで PDF 化して公開する情報公開もあります。そんなことよりちゃんと使える形にしましょうという

のが技術者の間でよく議論されることで、その中でも Linked Open Data は使いやすい形ですよと言われています。

Web の発明者であり、Linked Data の創始者でもあるティム・バーナーズ・リーさんはオープンデータについて、まずあなたのデータをオープンライセンスで Web 上に公開しましょう。URI を使うことで他のデータとリンクできます。一番いいのが LOD だよと言っています。

それでは LOD の技術普及が国内でどんな感じで進んでいるか、ここ数年の流れを紹介しますと、2011 年、「LOD チャレンジ」というのをわれわれが始めました。LOD の技術普及を目的としたオープンデータに関する日本初のコンテストで、以後毎年開催しています。

最初集まらなかったらどうしようと心配していたのですが、初年度の 11 年も 73 作品とそこそこ集まり、次の年 12 年は 205、13 年が 321、14 年は 245 と徐々に増えている。応募数から見ても広がってきていることが分かります。なお、15 年度も開催されます。賞金も出るし、学生奨励賞もあるので、ぜひ参加してください。

これと並行して日本国内でのオープンデータは思いのほか広がりができてきていて、われわれがコンテストを始めた翌年の 2012 年 7 月、「電子行政オープンデータ戦略」が政府のオフィシャルな文書として出ています。その中を見ると、基本原則の一つとして「機械判読可能な形式で公開すること」と、すでにこの時点から明記されている。

翌年 2013 年 5 月 24 日、「世界最先端 IT 国家創造」宣言（案）が出され、ここでもオープンデータの推進、ビッグデータの活用促進が言われています。翌月 2013 年 6 月 18 日には G8 サミット的首脳宣言で「オープンデータ憲章」に合意している。細かいところは後ほど見ていただければと思うんですが、原則、政府のデータをオープンデータとして公開すべきだとか、何年後にどれぐらい進んだか確認しようとか書かれていて、これも LOD 技術普及の追い風になると期待しています。

こういう流れの中で、2014 年 10 月から Data.go.jp という日本政府のオープンデータカタログサイトが本格運用開始になっています。13 年 6 月 18 日に G8 の「オープンデータ憲章」に合意しているの、その年のうちに日本政府としても

オープンデータを推進しているという姿勢を何か出さなければいけないということで13年12月に始めたのですが、予算の切れ目で14年4月に休止されてしまったという経緯があります。

ただ、このときにオープンデータはすごいなとみんなが確認したのは、オープンデータとして公開されているのはダウンロードしてよいので、ある人が止まる前に全部のデータをダウンロードして持っていた、その人が自分のURLでドメインを提供して、ミラーサイトがすぐ立ち上がった。半分冗談のようですが、オープンデータとしてデータを公開することがいかに有益かということが分かっていた話だと思えます。

そういう流れで次世代統計利用システムも2014年10月31日からAPI機能の本格運用が開始されていますし、特に経済産業省がオープンデータに力を入れていまして、LODを使った検索ができるサービスも提供しています。

(資料8) 通産省のOpen DATA METIのLODを活用したデータの可視化ワークショップも開催されています。細かくはこのサイトに行ってきたんですが、こんな感じでいろいろなことができるので、レポートを書くためにも学生さんには勉強してもらったらいいのではないかと思います。

2012年7月には産官学共同で「オープンデータ流通推進コンソーシアム」ができて、2年後の2014年10月14日に「オープン&ビッグデータ活用・地方創生推進機構」と名前が変わっています。オープンデータ活用とビッグデータの活用と別々に走っていたところが一つになったうえで、地方創生に活かそう。いま一つの運用として注目されているのは、各自治体レベルの地方のデータを地域活性化に活かそうということが強く言われているわけです。

総務省でも「オープンデータ戦略の推進」という専門ページがありまして、ここに行くと、いまの日本の政府関係の現状が全部まとめて見られるようになっていきますし、いろいろなプロジェクトも立ち上がっています。

われわれが「LODチャレンジ」を始めたころは日本の都市でオープンデータに取り組んでいたのは鯖江市、横浜市ぐらいでしたが、いまは地方創生の絡みもあって、150都市がオフィシャルに

オープンデータを公開しています。(資料9) これは鯖江市で早くからオープンデータの推進をされていたベンチャー系企業社長の福野さんがまとめたマップですが、いまこんな感じでオープンデータが進められているよということです。

ところで、「シビックテック」とか「コード・フォー・アメリカ」とか聞いたことある方いますか……。ないですね。アメリカではCode for Americaが有名ですが、ハッカーを集めて地域問題を解決するためのアプリケーションとかシステム開発をどんどんさせましょうという仕組みをつくったわけです。ハッカーというと攻撃した云々と言われますが、本来はIT技術を持った優秀な技術者という意味です。Code for Americaにならって、日本でもCode for Japanというのが立ち上がって、自分たちの住んでいるまちをよくするための活動をされています。

それをやるためには、使えるデータがないと何もできない。まずオープンデータをうまくつくって、それを上手に使うことで世の中のためになるようなアプリをつくらうと、いま全国で盛り上がっています。6月22日、その関係者がみんなが集まるような大きなイベントが横浜でありますし、東京では毎日曜のようにいろんなイベントがあります。後でネットに上げますので、ナマで実際に見たい方は一回行ってみられるとよいと思います。

まとめますと、最初は草の根的にわれわれがコンテストをやっていたんですけども、いつの間にかオープンデータ推進の大きな流れができて、自治体の取り組み、中央政府の取り組み、それに加えてシビックテックとかコード・フォーの取り組みとか、いろいろ混在してごっちゃになっているというのが現状です。

それではLODを使ってどんなことができるのかということですが、(資料10)これが「LODチャレンジ2013」の基盤技術部門優秀賞です。大阪市の橋が何年にできたかということもLODとしてつくって、こちらと連動して動くようなアプリをつくった。マッピング情報と年代情報があればSPARQLの検索だけでいろんなものの可視化ができるというようなアプリをつくって賞を取ったわけです。

(資料11) これは同じコンテストのチャレンジ

デー賞で、大阪での取り組みで、2年前から大阪市がオープンデータを始めた。国と協力して、このオープンデータが非常にいいとして公開ということをやった。

(資料12) それをしておくと、翌年ぐらいに、大阪市のオープンデータを使って、あらゆる必要な情報を Web とかスマホとかで簡単に検索できるようなアプリを実質1日のハッカソンでつくってしまった。

(資料13) これは去年の「LOD チャレンジ」のビジュアライゼーション部門優秀賞で、複数のデータを重ね合わせると何ができるかということで、警察署・交番の場所と犯罪の起こった場所を重ねてみると相関が見えますよという話です。

(資料14) これは学術的なもので、病気がどういう原因やメカニズムで起こるかというオープンデータと、病気の進み方についての DBpedia と、解剖などの 3D 画像をつなげて見えるようなシステムをつくっていますが、時間がなくて省略します。

このように、データとデータをつなぐというのは基本的な考え方で、それをどうやって見せるとどんな使い勝手ができるかというところで、いろいろなものが世の中でつくられている。それはデータを使う側からの観点ですが、逆にデータを出すほうから言うと、ただ単にデータを公開して見てもらうだけじゃなくて、いろいろなアプリケーションに応用してもらいたい。その両方で、いまいろいろなプロジェクトで LOD の公開が進められています。

(資料15) これは DBpedia の日本語版ですが、これを使ってもかなりいろんなデータが取れます。来月9日、DBpedia シンポジウムで DBpedia の活用特集をやるので、実際にどんな活用ができるか、興味がある方は来てください。

(資料16) ライフサイエンスのデータベースをつくって、みんなで公開して統合していこうというプロジェクトがずっと長く続いています。ここがすごいのは、海外とコネクションして、BioHakathon といって、1週間泊まり込みでデータをつくったりアプリをつくったりする英語でのイベントを毎年やっている。ことし8回目、長崎で開催します。それと別に、SPARQLthon といって、SPARQL を使って面白い検索とか情報交換も

毎月やっている。かなりアクティブなところで、日本で技術的に一番進んでいるのはこのコミュニティです。

(資料17) もう少し一般的なものは国立国会図書館の LOD です。図書館が持っているデータを LOD としてすでに公開しているのは、一つはどんな本があるかという書誌データです。典拠データというのは、図書館の本を分類するときに、これまでキーワードで人手で分類していたのが LOD 化されて、カテゴリーなどが見れるようになっていて、それがあれば、図書分類の関係性とかほかのデータの解析を組み合わせて何か見えてくる。たとえば NHK の番組情報とこのデータを組み合わせて番組情報の解析をしたりということもできると思います。

(資料18) J-GLOBAL Knowledge というのは、日本の科学技術振興機構 (JST) がデータをいっぱい持っている。データを内に持っているだけでは先細りだという危機感から、データを LOD として公開してどんどん使ってもらおうということで、このサイトをつい最近立ち上げて一般公開されています。いまは科学物質のデータだけが RDF で公開されていますが、ほかのデータを公開する準備も進められています。

以上をまとめますと、国内外ともに、すでに多くの LOD が公開されていて、使おうと思えば使える状態になっていることをまず知ってください。しかも、オープンデータの LOD 化が推進されるので、これからどんどん増えていくことは間違いない。現状は国内外で公開されているオープンデータの中で LOD の割合は数%以下と非常に数が少ないので、それが今後どう発展していくか、見どころでもあり、課題でもあります。

今後を展望しますと、われわれがコンテストを始めた 2011 年ごろはデータがなくて、アプリケーションをつくりたくてもつくれなかったんですが、いまいろいろなデータが公開されているので、それをどう組み合わせて活用するかというフェーズに入っています。それは趣味で「面白い」というレベルもあれば、ビジネスとして儲かるのではないかと企業が考える場合もありますが、いずれにしても有用なデータの組み合わせを見つけることで、そのためには LOD を正しくつなげることが重要です。その前に、ただデータを公開す

るのではなくて、つながるかたちでデータを公開することが大前提だということをぜひ知ってください。

いまビッグデータと LOD を組み合わせたデータ活用がさまざまな領域で模索されています。Linked Data, Linked Open Data を組み合わせることで、ただ統計的なデータを解析するのとは違った何か出てくるのではいか。そういう研究も走り

つつある状態です。

ホームページは誰でも簡単に公開できるようになりましたが、Linked Open Data を公開するのはまだまだ難しいところがあります。情報関係に携わる者としては、もっと手軽に Linked Open Data が公開できるような基盤技術やツールをつくることが大事だなあと考えているところです。

長くなりましたが、以上で終わります。

参考：

文中の（資料 1～18）については以下の URL を参照してください。

（資料 1）http://www.ted.com/talks/tim_bemers_lee_the_year_open_data_went_worldwide

（資料 2）<http://lod-cloud.net/>

（資料 3）<http://www.google.com/intl/bn/insideseach/features/search/knowledge.html>

（資料 4）<http://www.google.co.jp/search?q=村上春樹>

（資料 5）<http://www.w3.org/DesignIssues/LinkedData.html>

（資料 6）<http://ja.dbpedia.org/resource/大阪市>

（資料 7）<http://ja.dbpedia.org/sparql>

（資料 8）<http://datameti.go.jp/利活用事例>

（資料 9）<http://fukuno.jig.jp/2013/opendatamap>

（資料 10）<http://uedayou.net/SPARQLTimeliner/>

（資料 11）<http://lodosaka.hozo.jp/OsakaCityMapLOD.html>

（資料 12）<http://moyori.lodosaka.jp/>

（資料 13）<http://uedayou.net/osakacrimemap/>

（資料 14）<http://lodc.med-ontology.jp/>

（資料 15）<http://ja.dbpedia.org/>

（資料 16）<http://lifesciendb.jp/>

（資料 17）<http://www.ndl.go.jp/jp/aboutus/atandards/lod.html>

（資料 18）<https://stirdf.jglobal.jst.go.jp/>

TED Talk by Tim Berners Lee (1) 2010/2 ～LODの活用事例～



Sign In | Register

TED Ideas worth spreading

Talks TED Conferences TED Conversations About TED
 Speakers TEDx Events TED Community TED Blog
 Playlists (new) TED Prize TED-Ed (new) TED Initiatives

http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html

New TED Talks are released every weekday. Be the first to know!

Enter your email for TED updates Daily Weekly

Follow TED

TALKS | IN LESS THAN 6 MINUTES

ティム・バーナーズ＝リー「オープンデータとマッシュアップで変わる世界」

FILMED FEB 2010 • POSTED MAR 2010 • TED2010

494,797 Views 233

オープンデータの活用事例を紹介

- ・白人/黒人の住む家と水道管の整備状況の相関を見える化→裁判で勝利へ
- ・Where does my money go? (税金はどこへ行った?)

(英国) <http://wheredoesmymoneygo.org/>
 (横浜市) <http://spending.jp/>

そして白人が住む家と水道のある家との間に

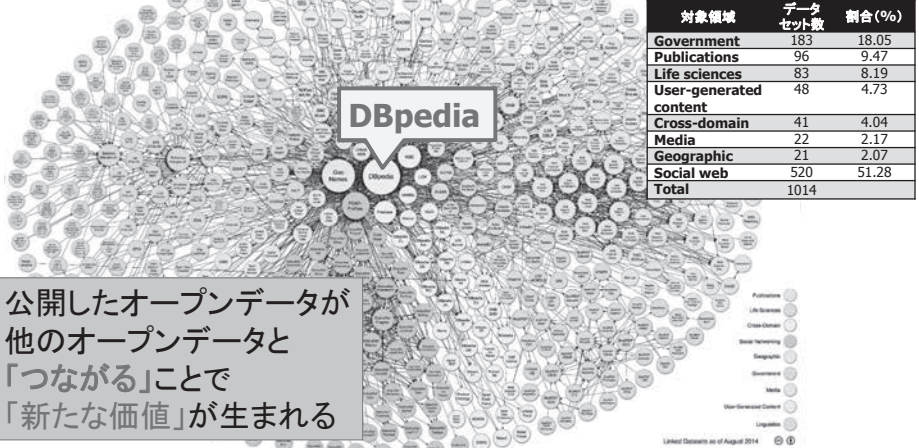
2015/6/18 1

既に公開・リンクされているLOD ～LODクラウド～



1つの丸が個別に公開されたDBを表す。

2014/08/30時点



公開したオープンデータが他のオープンデータと「つながる」ことで「新たな価値」が生まれる

Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

2015/6/18 2

まとめ①: LODの学術的/技術的背景



■ 従来のWebからSemantic Webへ

- 文書のWebから, データのWebへ
- 人間のためのWebから, 機械のためのWebへ
＝機械が「データの意味を処理できる」

■ Linked Dataの基本思想

- データを“つなげる”ことで新たな価値が生まれる
- Web上に公開された様々なデータを“つなげて”, 1つの大きなデータベースへとする

2015/6/18

3

Linked Dataの4つの条件



1. Use **URIs** as names for things
全てのモノやコトにURIをつけましょう
2. Use **HTTP URIs** so that people can look up those names.
それらのURIをhttp (Webブラウザと同じ方法) で参照 (アクセス) できるようにしましょう
3. When someone looks up a URI, **provide useful information**, using the standards (RDF*, SPARQL)
そのURIを参照したら, 役に立つ情報を提供するように
4. Include **links to other URIs**. so that they can discover more things.
多くのモノ・コトを発見できるように, 外部へのリンクを含めよう。



原文引用元 <http://www.w3.org/DesignIssues/LinkedData.html>
日本語訳参考 <http://www.slideshare.net/takeda/lod-5163454>

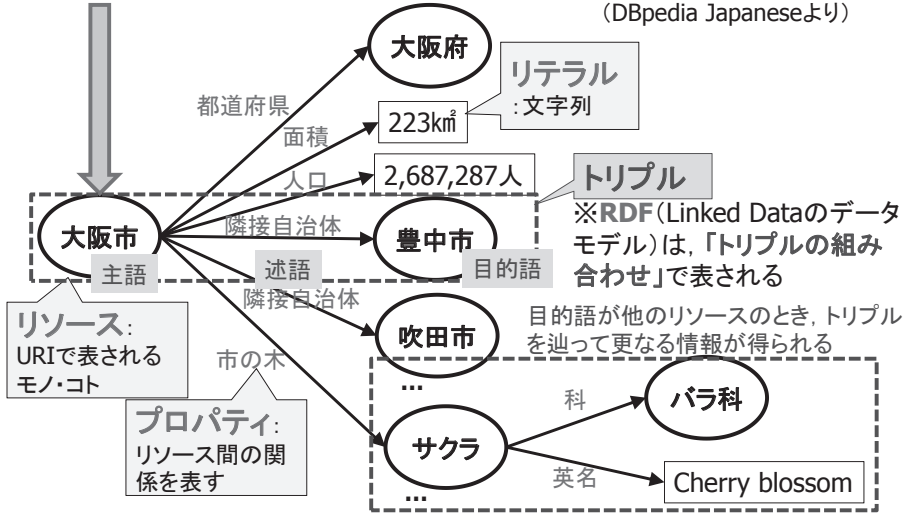
2015/6/18

4

RDFの例



<http://ja.dbpedia.org/resource/大阪市> というURIから得られる情報 (DBpedia Japaneseより)



2015/6/18

※実際のリソースとプロパティは、すべてURIで表される。

5

データを「つなげる」仕組み



- 「3つ組(トリプル)」により様々な構造のデータの“つながりを柔軟に表現”できる
 - Webサイトのリンクを辿るのと同様に、プロパティ(リンク)を辿ることで関連するデータの情報を辿ることが出来る。
 - RDFのプロパティは、“関係の意味”を定義できる。
 - cf. Webのハイパーリンクは単に“つながり”を表すのみ
- データ(リソース)をURIで表すことで、“外部のデータとつなげる”ことができる。
 - URIは、グローバルに一意的IDを表す。
 - WebサイトのURLと同じ仕組み。
 - cf. 単なる数字をIDとすると、異なるDBが同じIDを使っている可能性がある

■ Linked Data = 外部のデータとつながったデータ
≠ RDFフォーマットのデータ

2015/6/18

6