

行政のオープン化と Linked Open Data



国立情報学研究所 教授
総合研究大学院大学 教授
武田 英明

1 社会の変化と Linked Open Data

社会の変化に応じて行政の活動を変えるのは当然であろう。近年の社会の変化といえば、インターネットの普及であった。80年代に始まったインターネットの普及からの30年の社会変化は急激であった。インターネット普及と社会の変化についてはLessigの議論 [1] が参考になる。法学者であるLessigは人々の行動や社会秩序を規制するものとして、4つの要素を提示している。社会は法、慣習（規範）、市場、アーキテクチャという4つの異なるものの影響によって形作られているというものである。ここでいうアーキテクチャとは情報通信交通を含む社会の物理的・情動的仕組みを指している。技術の普及による社会変化とは、アーキテクチャの変化による社会の変化である。

インターネットの普及は80年代に始まっているが、ことに90年代に始まったWebの普及による変化のほうが現在の我々の社会への影響が大きい。Webは分散性、オープン性、経済性、簡便性という4つの特徴をもつコミュニケーション手段である。この4つのどれもがこれまでのコミュニケーション手段としてなかったもので、Webの導入は社会に強く影響を与えた。これらの4つの特徴は

そのまま90年代以降の我々の社会の特徴として埋め込まれるようになった。

Webの4つの特徴をもう少し詳しく説明する。分散性とは、どこにも中心となる場所がないということ、どこかに集中的に制御されるような構造ではないということである。オープン性は情報の出し手受け手双方において制約がないということである。経済性とはその利用において低廉であることで、ことにデータ量に依存しない仕組みであることが重要である。簡便性とは専門家でなくても誰でも情報提供や情報利用ができることである。

実際、我々の社会は情報の流通に関して着実に分散的でオープンで経済的で簡便になってきている。例えばいまあるレストランの情報が欲しいと思った時、レストラン自身が用意した情報を（分散性）、無料で（経済性）、簡単に（簡便性）、制約なく（オープン性）手に入るのは当然と思うであろう。しかし、これはWeb以前では当然ではなかった。経済性や簡便性は他のアーキテクチャの変化や市場の変化によるところもあるであろう。しかし、分散性とオープン性はWebからの導入により持ち込まれたものである。この社会の変化は我々の社会のアーキテクチャが変わらない限り続くであろう。

またインターネット/Webが社会のアーキテクチャになったゆえのもう一つの変化はデータの重

要性である。インターネット以前の社会において、情報のやり取り・処理は基本的に人間の仕事であった。インターネット/Web時代においては、情報は人間が読み使うだけではない。人間が直接読むのではなくてコンピュータが処理するための情報、すなわちデータが重要になった。行政における情報も同様で、人間向けの情報だけでなく、コンピュータのための情報（データ）も重要である。

2

データのWebとしての Linked Open Data (LOD)

Linked Open Data (LOD) はいわばWebのデータ版で、Webの特徴そのまま、それをデータの流通に使うというものである。通常のWebは主に文書のネットワークを作っており、「文書のWeb（蜘蛛の巣）」である。異なる組織や個人が異なる場所で公開した文書（Webページ）がお互いにリンクし合っている。一方、LODは「データのWeb（蜘蛛の巣）」である。異なる組織や個人が異なる場所で公開するデータがお互いにリンクし合うというのがLODである。これまでのデータの公開はたとえWeb上に公開されていたとしても、それぞれが孤立しており、相互につながることはできなかった。通常のWebとLODで違う点は、Webではリンクは1種類であるが、LODではリンクにラベルがあり、さまざまな属性をつけたリンクをつくることができる。

LODはWebと同じようにグローバルに共有するデータ空間である。すなわち、データが世界のどこのだれが管理しているデータセット（データベース）に含まれているかを意識することなく、アクセスしたり、リンクす

ることができる。これまであったデータセットとデータセットの間の障壁は存在しない。個々のデータセットはお互いにリンクし合うことで、一つのグローバルなデータセットの一部となるわけである。

3

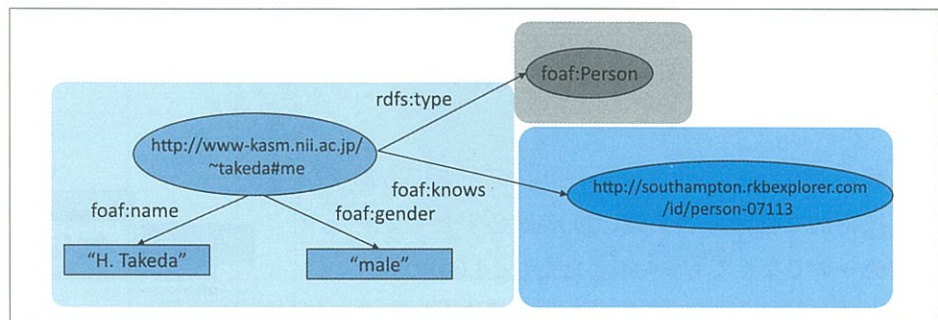
LODの仕組み

図表1はLODで表現した例である。丸はURI（URLを一般化したもの）、四角はリテラル（文字列）を示している。その間をラベル付きのリンクが結んでいる。このラベルがリンクの性質を示している。また大きな枠は異なるサイトにあるということを示している。この例では、二人の人物のデータが異なるサイトにあり、繋がっていることを示している。このデータはRDF（Resource Description Framework）というモデルを使って記述される。RDFでは全てのデータは矢印にあたるような主語—述語—目的語という簡単な形式で書かれる。（図表2参照）

たとえば「AさんはBさんを知っている」という関係は「A knows B」という3つ組で表現される。このとき表現したい事物（たとえば、前述の例ではA、B）には個別のURIを与える。ここでは個々の事物にURIを振ることで、URIが世界中でユニークなIDとして使えることが重要である。

さらにデータを記述する様式が決まっている場合、その様式をスキーマとして別途定義する。

図表1 LODによる表現



図表2 RDF (Turtle) による記述

```
<http://www-kasm.nii.ac.jp/~takeda#me> rdfs:type foaf:
Person .
<http://www-kasm.nii.ac.jp/~takeda#me> foaf:name "H.
Takeda" .
<http://www-kasm.nii.ac.jp/~takeda#me> foaf:gender
"male" .
<http://www-kasm.nii.ac.jp/~takeda#me> foaf:knows
<http://southampton.rkbexplorer.com/id/person07113> .
```

RDFを拡張したRDF Schema (RDFS) という言語では、データ記述の様式をクラスとプロパティの組み合わせで定義する。クラスはある事物や事象を示し、プロパティはクラスがもつ属性やクラス間の関係の型を示す。RDFSはRDFの拡張であるので、クラスやプロパティの定義もURIを与えて、参照可能にする。個別の事物や事象はこれらのクラスのインスタンスとして記述される。このようなスキーマ定義はデータセットごとに用意する必要はない。すでに他で定義されたスキーマがあるときは、それを使うことができる。

URIとRDFを使うことでデータはデータがどこに含まれているといったことを気にせずにデータを相互につなげることができる。これは新しいデータの世界である。Tim Berners-LeeはこのLODを普及させるために4つの原則を提唱している。

1. 事物をURIを使って名前付けしよう
2. 名前の参照がHTTP URIでできるようにしよう
3. URIを参照したときに関連情報が手に入るようにしよう
4. 外部へのリンクも含めよう

この原則にもとづくとデータセットは相互につながり、データセットのネットワークができる。

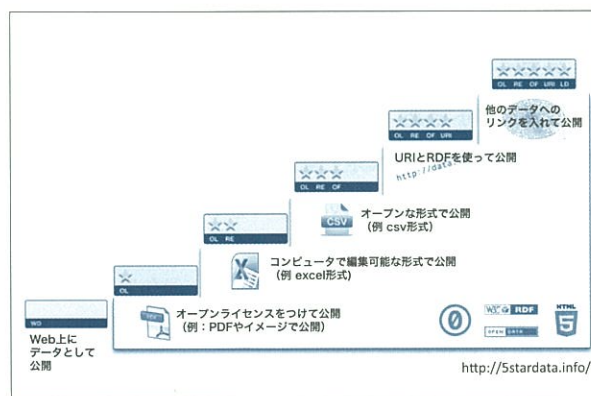
4 オープンデータとLOD

名前の通りLODは通常はオープンデータでもある。技術的な面から言えば、LODは必ずしもオープンデータでなくてもよい^{注1)}。ただし、LODの利点を生かすにはオープンであることは望ましい。こ

れはHTMLをインターネットに使うのかイントラネットに使うかの違いと同じである。HTMLをイントラネット用、すなわち特定のサイト内に閉じたHTMLファイル群を作ってWebを運用することはなんら問題ない。しかし、当然イントラネット内のページは外からリンクをつけることはできないし、検索にひっかかることもない。すなわちHTMLが持ついくつかの機能は活かされない。LODも同様で、閉じた環境でも使えるが、オープンな環境で使ったほうがフルに機能を発揮できる。オープンデータのほうから見ると、LODはより高次の機能をもったオープンデータの形式ということになる。そこで、Tim Berners-Leeはオープンデータ構築の視点からLODを図表3のように位置付けている。この図はオープンデータの五つ星 (Five Star Open Data) と呼ばれている。

単にWebにデータを構築するのは星0と考える。その上で、オープンライセンスを付与して公開するものを星1とみなす。ここで初めてオープンデータになる^{注2)}。この時点ではどのような形式のデータでもよい。次に“機械可読可能な”形式で公開することを星2とみなす。ここで機械可読可能な形式とは、コンピュータのソフトウェアで容易に編集が可能なような形式である。例えば、表を画像としたものは表の中の文字をコピーしたり変更したりすることができないので、機械可読可能とはいわない。さらにこのデータ形式がオープンな形式であるとき、星3と呼ぶ。オープンなデー

図表3 5つ星オープンデータ



タ形式とは、データ形式自体が公開されていて、誰でもこの形式を使ってよいものである。そうでない、特定のソフトウェアでないと読めない形式は含まれない。ここまでが“一般の”オープンデータのレベルである。

さらにデータ形式がRDFであるとき星4とみなす。RDFで書くことで、異なるデータセットを容易に統合でき、外からも容易にアクセス可能になる。さらに外部へのリンクを含むようなRDFで書いたときを星5とみなす。これはまさにLODである。この段階でデータは相互につながりあい、グローバルなデータ空間の一部として活用可能になる。

5

LODによりつながるデータの世界

現在、上記の原則に基づいたLODが多数公開されている。図表4に示すのは、世界中のLODのネットワークを示す図で、これをLOD Cloudと呼んでいる。2014年8月の時点での図で、500以上のデータセット（データベース）が含まれている。丸が個別のデータセットを示し、データセットとデータセットを結んでいる線は、データセット内のデータ同士にリンクがあることを示している。中心にあるのはDBpediaというもので、Wikipediaの情報をLOD化したデータセットである。Wikipediaは百科事典なので、さまざまな分野の項目が含まれている。このため、他の多くのデータセットと結びつきやすい。このため、DBpediaはこの図の中心にある。

図表4ではドメインごとにデータセットが色分けされている。右上の部分は出版物に関連するデータセットで、図書館の書籍書誌データやジャーナル論文の書誌情報といったものが含まれている。右下

はバイオサイエンスなどを含むライフサイエンスに関するデータセットである。左端は政府関係のデータセット、すなわち多くは行政に関するデータセットである。ここにはアメリカの政府データポータルであるdata.govやイギリスのポータルdata.gov.ukのデータのうち、LODのものが含まれている。このLOD Cloud図は2008年以降順次更新されているが、この数年で最も増えているがこの政府関係のデータセットである。これは各国のオープンデータ政策にリンクしており、オープンデータとして公開されるもののある部分がLODとして公開されるようになり、その結果、政府関係のLODデータセットが増えている。

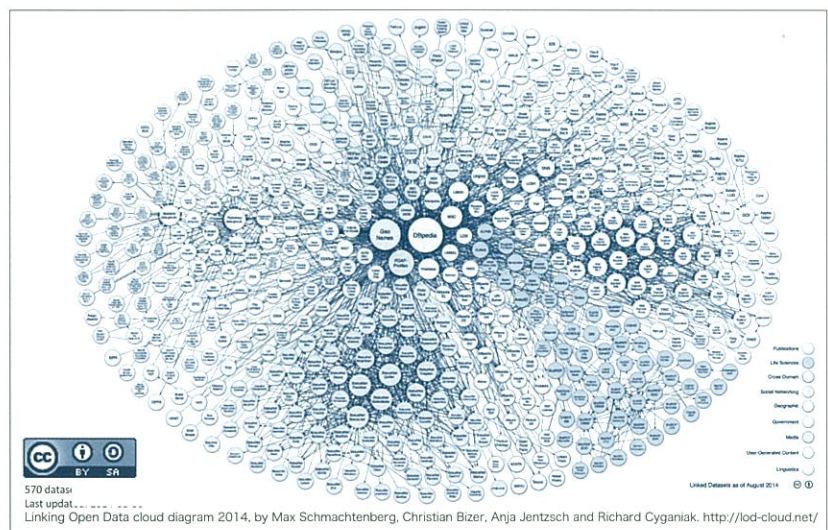
6

世界の行政関係のLODの現状

各国のオープンデータ政策についてはすでに他の記事で触れられているので、ここではLODとの関係について述べる。

アメリカ合衆国では data.govを2009年より開設している。その開設当初からSemantic Webという項目があり、これはデータのLOD化についての項目で、data.govで公開されたデータの一部を大

図表4 LOD Cloud



学のチームが中心となってLOD化して公開していた。

イギリスではアメリカに続いて2010年よりdata.gov.ukを開設したが、ここでもトップページのメニューにLinked Dataというものが用意され、当初からLOD化を進めている。イギリスの場合、官民共同でオープンデータを推進する組織 Open Data Institute (ODI) を2012年に設立した。ODIのミッションは行政のオープンデータ化における技術的支援、コミュニティ支援、イノベーション支援であるが、その創設者の一人はLODを推進するTim Berners-Leeであり、LODはその技術的射程に含められている。

イギリスの行政データのオープン化におけるLODをもう少し詳しく見てみよう。早くからオープンデータに積極的に取り組んできている陸地測量部 (Ordnance Survey) では地理関係の大規模データセットを2010年よりLODにより公開している。また英国ではコミュニティ・地方自治省 (DCLG) が多様な地方行政データをLODとして公開しており、現時点で215データセットが登録されている。このDCLGやグレーター・マンチェスター、スコットランド政府等は研究プロジェクトの成果を用いて構築されたSwirrl社のLinked Data公開プラットフォームPublishMyDataを用いて統計データなどをLODとして公開している。

7

国内における行政データのLOD化の現状

国内においても、行政データのLOD化は行政におけるオープンデータの推進の一部として行われてきた。行政データのオープンデータ化の立ち上げ時からLOD関連の団体が協力していたため、いくつかの先駆的事例がみられる。

たとえば鯖江市が位置情報を含む施設等の情報をLOD化して提供し、それを地図表示アプリの中

で利用するというを行っている。横浜市では施設等の情報だけでなく、アート関連のイベント情報もLODとして提供しており、これも地図アプリなどで利用されている。横浜市金沢区では子育て情報を提供するサイトの情報をつくるため、区役所各部門でのデータをRDF化したのち統合する仕組みを実装している。これはRDFがデータ統合手段として優れているためである。国レベルでは、統計センターは都道府県コードをLOD化して提供している。

国内には約200の自治体がオープンデータをすでに公開している^{※3}。ただし、LODとしてデータを公開する自治体は数的にはまだそれほど多くない。秋山ら [2] によれば、RDFを公開している自治体は51、データセット数で約1,200件である。(図表5参照)

図表5 RDF を公開している自治体 [2]

1	北海道室蘭市	27	長野県駒ヶ根市
2	北海道森町	28	長野県中野市
3	北海道八雲町	29	長野県塩尻市
4	宮城県石巻市	30	長野県軽井沢町
5	秋田県横手市	31	静岡県静岡市
6	福島県会津若松市	32	静岡県三島市
7	埼玉県川口市	33	静岡県島田市
8	埼玉県和光市	34	静岡県磐田市
9	埼玉県北本市	35	静岡県掛川市
10	千葉県流山市	36	静岡県裾野市
11	千葉県我孫子市	37	静岡県御前崎市
12	東京都品川区	38	愛知県長久手市
13	東京都杉並区	39	三重県津市
14	神奈川県鎌倉市	40	滋賀県大津市
15	新潟県三条市	41	大阪府高槻市
16	新潟県十日町市	42	大阪府枚方市
17	新潟県見附市	43	兵庫県神戸市
18	新潟県糸魚川市	44	鳥取県
19	富山県砺波市	45	島根県松江市
20	福井県	46	岡山県玉野市
21	福井県福井市	47	広島県呉市
22	福井県敦賀市	48	山口県宇部市
23	福井県鯖江市	49	山口県周南市
24	福井県越前市	50	香川県高松市
25	長野県上田市	51	熊本県菊池市
26	長野県須坂市		

8

自治体における
LODの普及にむけて

RDFで公開することにより、データはより使いやすくなる。しかし、csvといった形式に比べてRDFで公開することは容易ではない。ただ、近年RDF化、LOD化を支援するツール・サービスがいくつも作られている。商用を含むいくつかのLOD公開支援サービスが登場して、技術的知識がなくても保有するデータをLODとして公開できる環境が整ってきた。LinkData.orgは早くから簡便なLOD化ツールを提供してきたが、jig.jp社ではオープンデータプラットフォームという有料サービスを2014年6月より開始し、地方自治体などが採用している。この他、Datashelf（インフォランジ社）などもある。これらを使うことで手元にテーブル型データ（excelデータなど）を簡単にRDF化、LOD化することができる。

また、LODとしての質を高めることも、より使われるデータをつくるために重要である。このためには秋山ら [2] が提案したLOD 実現のための11の基準が参考になる（図表6参照）。これは3章で挙げたLODの4つの原則をより具体化したもので、どのようなデータをつくれればよいかを決めるときに参考になるであろう。

図表6 LOD実現のための11の基準 [2]

	基準	内容
1	機械可読	機械判読可能であること
2	IRI付与	事物の名前としてIRIが付与されていること
3	HTTP IRIの利用	名前を参照できるよう、HTTP IRIを用いていること
4	共通の語彙の利用	共通の語彙を利用していること
5	語彙の正確な利用	語彙の利用が正確であること
6	IRIのドメイン	リソースのドメインがリソースを管理する自治体の所有するドメインであること
7	IRIの参照	主語IRIの参照する先が存在すること
8	HTTPコンテンツネゴシエーション	HTTPコンテンツネゴシエーションにより、要求に応じた形式で情報提供をしていること
9	アウトバウンドリンク	外部IRIへのリンクを含むこと
10	実在するIRIへのアウトバウンドリンク	参照解決可能な外部IRIへのリンクを含むこと
11	RDFリソースへのアウトバウンドリンク	参照解決可能なRDFリソース IRIへのリンクを含むこと

9

おわりに

本稿では、LODの基礎的な考え方から自治体での取り組みまでを紹介した。LODはデータの活用においては大きな可能性をもっているが、必ずしも平易ではない。先に述べたようにLOD化を支援するツールが登場し、コストは低減している。それでも一定のコストがかかる。他から公開されるデータと合わせると有用なデータ、他のデータとつながっているデータなど、LODに適したデータを優先して進めていくとよいだろう。

【参考文献】

- [1] Lessig, Lawrence. Code and other laws of cyberspace. Vol.3. New York: Basic books, 1999. (山形 浩生, 柏木 亮二 (訳, CODE—インターネットの合法・違法・プライバシー, 翔泳社, 2001)
- [2] 秋山梓, 加藤文彦, 小出誠二, 海沼靖夫, 自治体が公開しているRDFの現状と課題, 人工知能学会研究会資料, 2016.

【注】

- 1.この点を考慮して、Linked Open DataではなくLinked Dataと称することもある。名前の違いだけで指しているものは同一である。
- 2.オープンデータとは単にデータが公開されていることを示すのではない。加工や再配布を含めて使えることが許されたデータをオープンデータと呼び、そのような規約を用意したものをオープンライセンスと呼ぶ。オープンライセンスの1例はCreative Commons CC-BY である。
- 3.<http://fukuno.jig.jp/2013/opendatamap>