

公共データを民間へ開放する 「オープンデータ」の最新動向と今後の展望

長谷川 大祐
Hasegawa Daisuke

浅野 優
Asano Yu

近年、政府や自治体などが保有する公共データを二次利用可能な形で公開する「オープンデータ」の取り組みが、国内外で活発に推進されている。日本政府は、IT分野の戦略「世界最先端IT国家創造宣言」において、2015年度末までに他の先進国と同水準の公開内容を達成することを目標に掲げ、オープンデータを推進している。

日立グループは、公共機関のオープンデータの実現に必要なとなるプロセス全体をカバーし、オープンデータ推進に係る計画策定から、データの整備、公開、利活用までの一連の活動を支援する「オープンデータソリューション」を提供するとともに、データの活用性を向上させる次世代技術の研究開発に取り組んでいる。

1. はじめに

「オープンデータ」とは、政府や自治体などが保有する公共性の高いデータをインターネット上に公開し、一般利用者が利用できるようにしたデータにすること、およびそれを推進する活動である。オープンデータの取り組みにおいては、従前の「情報公開」とは異なり、二次利用しやすいように使用ライセンスやデータ形式に配慮した公開がなされる。わが国では2012年の「電子行政オープンデータ戦略」を端緒に、政府としての取り組みが始まった。2013年の「世界最先端IT国家創造宣言」において、めざすべき社会を実現するための取り組みとして取り上げられるなど、日本政府はオープンデータを積極的に推進しており、市民社会の盛り上がりも追い風に、その取り組みは一層の拡大を見せている。

本稿では、オープンデータに関する国内外動向を官民双方について紹介するとともに、オープンデータ推進を支援する日立グループの取り組みや、データの活用性を向上させる先進技術、今後のオープンデータの展望について述べる。

2. オープンデータの動向と国内外の取り組み

2.1 国内外のオープンデータ公開の動向

2000年代初頭、政府の透明性向上を目的に英国から始まったオープンデータの取り組みは世界各国に広がってお

り、すでに世界45か国で政府のデータカタログサイトが開設されている¹⁾。わが国も2013年6月に採択した「G8オープンデータ憲章」の内容を踏まえ、政府のデータカタログサイト「DATA.GO.JP」を2014年10月に開設した。中央省庁の保有する統計情報、地理空間情報などの利活用ニーズの高い公共データがオープンデータ化され、「DATA.GO.JP」を介して横断的な検索・収集が可能となった。

自治体に目を向けると、先進的な自治体において、福祉、観光、行政支出などの地域課題への対応に資するデータを中心に、政府の取り組みに先駆けオープンデータを提供してきた。さらに、福井県では、県内の市町村でデータのフォーマットをそろえて公開するなど、県内市町村を横断したデータ活用の推進をにらんだ取り組みも見受けられる。オープンデータに取り組む自治体はすでに100を超え(2015年6月末時点)、加速度的に増加しており、今後も政府の普及活動によるさらなる拡大が見込まれる(図1参照)。

一方、オープンデータ先進国として有名な米国および英国においては、政府のデータカタログサイトにおいてそれぞれ14万1,218件、2万5,664件のデータセットが公開されている(2015年6月末時点)^{2), 3)}。また、米国では2014年5月に「データ法(Data Act)」が成立しており、米国政府機関の財務・支出データについて指定のフォーマット・項目名で公開することを義務付けている。

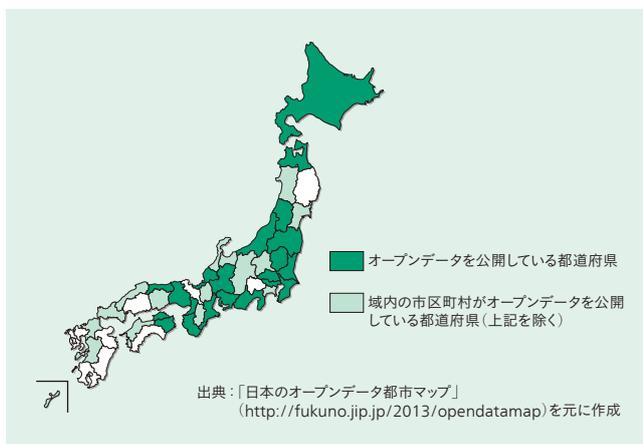


図1 自治体におけるオープンデータの取り組みの広がり

オープンデータをすでに公開している自治体を地図上に示す。

2.2 オープンデータ公開の効果

ここまで述べてきたように、国内外でオープンデータの取り組みが盛んに推進されている。政府や自治体などがオープンデータに取り組む意義としては、以下のような効果が挙げられる。

(1) 行政の透明性向上

政府や自治体は予算データなどの行政活動に係るデータを公開することで、納税者である国民への説明責任を果たすことができ、国民の行政への信頼性向上や、行政への参加意欲向上などにつながる。

(2) 行政サービスの効率化

政府や自治体は観測データや国民生活に関するデータを、加工しやすい形で公開する。民間事業者がこれらのデータを活用し、国民のニーズやユーザビリティを考慮したアプリやサービスを開発・提供することで、情報提供や施策周知に係る行政側の業務が一部代替される。

(3) 経済の活性化

政府や自治体は統計や地図情報など、民間事業者などでは独自入手が困難なデータを公開する。民間事業者がこれらのデータを活用し、新ビジネスの開発、マーケティング活動の効率化、業務の効率化などを実現することで、経済振興に結びつく。

また、オープンデータ公開に取り組むのは、行政機関だけではない。例えば、東京地下鉄株式会社(東京メトロ)は2014年に、リアルタイムの列車運行情報などをオープンデータとして公開し、このデータを用いた「オープンデータ活用コンテスト」を開催した。保有データをオープンデータとして公開することで、多くの開発者がコンテストに参加し、最終的に281件のアプリが生み出された⁴⁾。

2.3 オープンデータを活用したビジネス事例

オープンデータの先進諸国では、民間領域においてオー

表1 | オープンデータ活用事例

海外のオープンデータ利活用の一例を示す。

カテゴリ	サービス・アプリ概要	オープンデータの利用方法	収益方法
不動産	不動産仲介業者向け不動産物件情報提供サービス	地域の環境・労働・教育・気候などの不動産価値に資する情報を収集	●サービス利用料
企業情報	103か国、約8,500万社の企業情報(登記住所、財務情報など)データベース	登記情報をデータベース上に収集、企業サイトのクロールも実施	●データ販売 ●サービス利用料(API使用料)
金融	不正利用の疑いがあるクレジットカード取り引きを利用者に通知するモバイルアプリ	過去の不正利用・苦情データから、不正検知のアルゴリズムを導出	●サービス利用料
交通	短時間かつ快適な(混雑率が低いなど)移動ルートを検索するモバイルアプリ	ルート検索時に交通事業者の運行情報を参照、混雑率などはアプリ利用者が投稿	●広告掲載料

注：略語説明 API (Application Programming Interface)

ペンデータを活用したビジネス事例が多数生まれており、その一例を表1に示す。不動産の価値判断への利用や、企業の信頼性評価への利用などがある。

3. 日立グループのオープンデータソリューション

3.1 オープンデータ公開までの取り組みのステップ

政府や自治体などの公共機関がオープンデータ公開へ取り組むにあたっては、保有データの把握や、データ公開環境の整備などが必要となる。公開までの取り組みのステップについては、一般社団法人オープン&ビッグデータ活用・地方創生推進機構が公開している「オープンデータガイド第1版」(図2参照)や、2015年2月に内閣官房から公開された「オープンデータをはじめよう～地方公共団体のための最初の手引書～」に詳しく説明されている。

「オープンデータガイド第1版」において、「現状把握」のステップでは、組織内で保有しているデータの種類や分量、管理部署などを把握するために、データの棚卸しを実施することが望ましいとされている。また、「公開作業」のステップでは、データカタログサイトなどのデータ公開環境の整備が、「公開・運用」のステップでは、継続的なデータの公開などが求められる。

しかし、行政機関が上述のサービスを整備するには、各機関が保持する大量データの整備や公開に係る業務負担の増加が想定される。それらを軽減するためには、オープンデータの公開環境を整備する支援が求められる。2014年7月に日立グループでは、これまでのオープンデータに関する研究・開発や事業実績を基に、「オープンデータソリューション」の提供を開始した。次節で、「オープンデータソリューション」について説明する。

3.2 オープンデータソリューションの特長

日立グループが提供する「オープンデータソリューショ

オープンデータ作成・公開の6ステップ

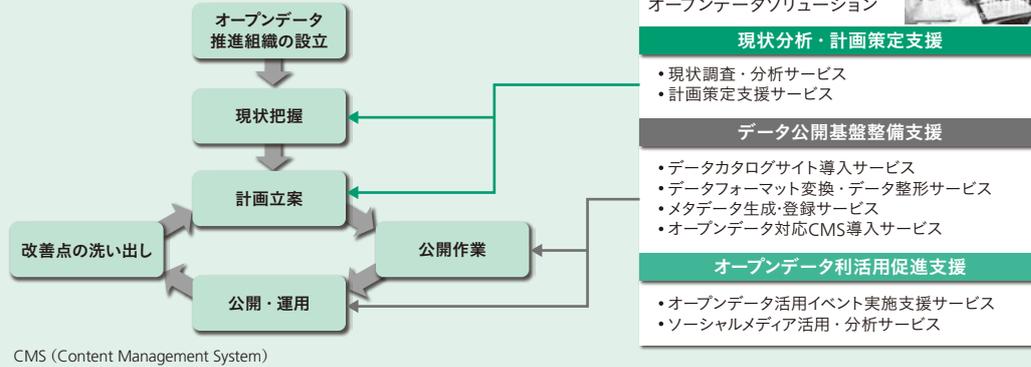


図2 | オープンデータ公開のステップと日立グループのソリューション

左にオープンデータ公開のステップを、右に日立グループのオープンデータソリューションのメニュー一覧を示す。また、それらの対応関係を併記する。

ン」は、公共機関のオープンデータの実現に必要なプロセス全体をカバーし、計画策定からデータの整備、公開、利活用までの一連の活動を支援するものである。「オープンデータソリューション」では、8つのサービスを提供している(図2参照)。以下に、主なサービスメニューについて紹介する。

(1) 現状調査・分析サービス

オープンデータを始めるにあたり、どのデータを公開すればよいか最初の課題となる。公開するデータを決めるには、まず組織内で保有されているデータの種類や量、形式、公開可否などの定量的な把握が必要となる。そのうえで、利用者(市民・企業)のニーズの高いデータを、利用しやすい形式・方法で公開する必要がある。

現状調査・分析サービスでは、保有データの棚卸し調査により情報資産の可視化を行うとともに、ニーズ分析の結果を組み合わせて、公開するデータの優先順位づけや公開方法などの計画を立案する。また、推進にあたっての制度的・技術的課題の抽出および対策を検討し、取り組み事項の整理を行う。

(2) データカタログサイト導入サービス

オープンデータを公開する際は、大量のデータの中から利用者が必要とする情報を容易に見つけられるようデータカタログサイトを構築することが一般的である。データ提供者はデータカタログサイトにオープンデータのメタデータ(データの書誌情報)を登録する。データ利用者はデータカタログサイトにアクセスすることで、キーワードで公開データを検索することや、公開データをダウンロードする前に、データの概要から利用したいデータか否かを判断することが可能となる。

データカタログサイト導入サービスでは、米国および英国政府で採用されている「CKAN」というオープンソースソフトウェアを用いたデータカタログサイトを構築・運用

するサービスを提供している。

内閣官房における適用事例では、日本政府のデータカタログサイト「DATA.GO.JP」を構築し、2014年10月から運用を開始した。現在、「DATA.GO.JP」では、1万3,038件(2015年6月末時点)のデータセットが公開されている。

(3) オープンデータ対応CMS導入サービス

データカタログサイトへのデータ登録にあたっては、職員によるメタデータ作成や登録の手間が発生する。職員の業務フローを考慮すると、ホームページへのコンテンツ登録と並行して、データカタログサイトへのメタデータの登録作業を行えることが望ましい。

オープンデータ対応CMS(Content Management System)導入サービスでは、自治体のホームページ運営を支援するCMS「4Uweb(フォーユーウェブ)/CMS」の「オープンデータカタログサイト自動連携」機能により、ホームページへのコンテンツ登録と並行して、データカタログサイトへの登録が行える。

4. オープンデータの先進技術

4.1 より二次利用しやすいデータ形式での公開

2章でビジネス事例を紹介したとおり、ビジネスにオープンデータを活用する動きが広がりを見せている。オープンデータの活用を促進するには、利用者が二次利用しやすい形式でデータを公開することが重要と考えられる。そのため、先進的な国では、機械(プログラム)がより処理しやすいデータ形式であるLOD(Linked Open Data)の提供が推進されている。

LODは「RDF(Resource Description Framework)というデータ形式で、機械が意味を理解して処理できる」、「データにユニークな識別子URI(Uniform Resource Identifier)が付与され、外部のデータとの参照関係が定義されている」という性質を持つオープンデータである。

LODを活用することで、組織の内外を問わず、データの横断的な検索や統合的な分析などが容易に行えるようになる。

4.2 統計表のLOD化技術

行政の保有するデータで、すでに多方面で活用されているものに統計データがある。ここでは、日立グループの技術の一端として、統計表を効率的にLODに変換可能な技術を紹介する。

統計表をLOD化するには、上述のように、CSV (Comma-separated Values) やXLSなどの表形式のデータをRDFに変換し、データにURIを付与する必要がある。RDFでは、データを主語、述語、目的語の3つの要素(トリプル)で表す。例えば、「日本の人口は1.273億人である」というデータは、「日本」を主語、「人口」を述語、「1.273億人」を目的語とするトリプルで表される。その際、「日本」、「人口」の各語を、ユニークな識別子URIを用いて表す。URIは、用語の表記にかかわらず、同一概念には同一のものを使用する。ここでURIとして、外部の組織がすでに使用しているものを用いることで外部のLODと連携させることができ、利用者が効率的に検索、活用できるようになる。

統計表は、数十万個以上の値を含む場合もあるため、人手でLODに変換するには手間がかかる。そのため、LODへの変換を支援するツールを活用することが考えられる。しかし、従来のツールは、特定の構造のデータしか変換することができず、多種多様な表構造がある統計表を変換するには、データ構造を変換する手間が発生する。したがって、いかに手間なくLODへ変換させるかが課題である。

この課題に対し、日立グループは、統計表を効率的にLODに変換するためのテンプレートと、それをLODに自動変換するプログラムを整備し、LOD変換支援ツールを開発した⁵⁾、⁶⁾。本テンプレートは、多種の構造の統計表に適用可能であり、データ構造を変換する作業量を減らすことができる。また、各語に対応する外部のURIを推薦する機能を持ち、効率的に外部のLODと連携させることができる。実際に、政府がオープンデータとして公開する6つの統計表に対して本技術を適用した結果、数百万個以上の値を含む統計表をLODに変換でき、そのデータが容易に検索、連携できることを確認した。

5. オープンデータの今後の展望

これまで述べてきたように、政府や自治体などでは、先進国と同水準のオープンデータ環境の実現が整備されつつある。しかし、オープンデータの取り組みを通じた経済の

活性化や住民サービスの高度化は、十分には達成されていないのが現状である。

オープンデータがより活用される社会を実現するには、(1) 公共機関によるデータ提供、(2) 民間事業者などによるサービス提供、(3) 市民・企業などからの公共機関に対するフィードバック、という3ステップを循環させ、相乗的に発展させることが必要であると考えられる。

日立グループは、オープンデータに取り組みたい行政機関やオープンデータを活用したい民間事業者に対して、その課題の整理から解決策の策定、実行に至るまで、先述のソリューションや技術を基に、オープンデータの公開および活用を支援している。また、顧客の既存業務の効率化や事業サービスの向上を図るべく、さらなるソリューション開発や技術研究を推進している。

6. おわりに

ここでは、オープンデータの動向と国内外の取り組み、データ公開側の課題を解決する日立グループのソリューションや技術、および、オープンデータの展望について述べた。日立グループは、政府や自治体などとのオープンデータを通じた社会イノベーションや、民間企業とのビジネス応用での協創による関連製品・サービスの実現をめざしている。

参考文献など

- 1) OPEN GOVERNMENT, DATA.GOV (2015.6)
<http://www.data.gov/open-gov/>
- 2) Search for a Dataset, DATA.GOV (2015.6)
<http://catalog.data.gov/dataset>
- 3) Datasets, DATA.GOV.UK (2015.6)
<http://data.gov.uk/data/search>
- 4) 東京メトロニュースリリース「東京メトロオープンデータの取組」(2015.3),
http://www.tokymetro.jp/news/2015/article_pdf/metroNews20150327_36-194.pdf
- 5) 浅野, 外: 統計データのRDF化のためのテンプレート, 第12回情報科学技術フォーラム講演論文集, 12 (2), F-034 (2013.8)
http://ci.nii.ac.jp/vol_issue/hels/AA1242354X_ja.html
- 6) Asano, et al.: A Template for Handling Statistical Data in RDF, Second International Workshop on Semantic Statistics (SemStats 2014) (2014.10)

執筆者紹介



長谷川 大祐
日立製作所 情報・通信システム社 公共システム事業部
公共戦略企画部 所属
現在、オープンデータ推進に関するビジネス企画に従事



浅野 優
日立製作所 研究開発グループ システムイノベーションセンタ
知能情報研究部 所属
現在、オープンデータ推進に向けた基盤技術の研究開発に従事
博士(情報科学)
情報処理学会会員, 人工知能学会会員, LODチャレンジ実行委員