

公共データの民間開放の 最新動向と今後の展望

—オープンデータ—

Latest Trends and Future Prospects for Openness in Public Sector Information

坂倉 芳崇 浅野 優 小池 博
Sakakura Yoshimune Asano Yu Koike Hiroshi

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。内閣官房は、高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部）に電子行政オープンデータ実務者会議を設置し、2013年6月に「電子行政オープンデータ推進のためのロードマップ」を発表するなど、取り組みを加速している。その目的は、府省・自治体活動の「透明性の向上・信頼性の向上」、「国民参加・官民協働の推進」、「経済の活性化・行政の効率化」にある。日立グループは、関連活動に積極的に参画し、公開する側と利用する側に向けたソリューションの整備を進めている。

1. はじめに

公共データの「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、一般利用者が自由に利用できるようにしたデータ、およびそれを推進する活動である。政府や自治体などの公共データの公開は、政府活動に関するデータの公開（透明性）と、官民共同・国民参画（双方向）の考えに立脚した「オープニングガバメント」から始まった。近年は、これらに「経済活性化視点」が追加され、二次利用しやすいライセンスとデータ形式での公開の視点が強くなり、オープンデータという概念に変わってきている。

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取り組み、および、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 オープンデータ化の取り組みと利活用事例

オープンデータ化の取り組みは、従来のWebサイトでのデータ公開に加え、以下の活動から成る（図1参照）。

(1) 表やグラフの元になっている未公開の数値データ（ロ

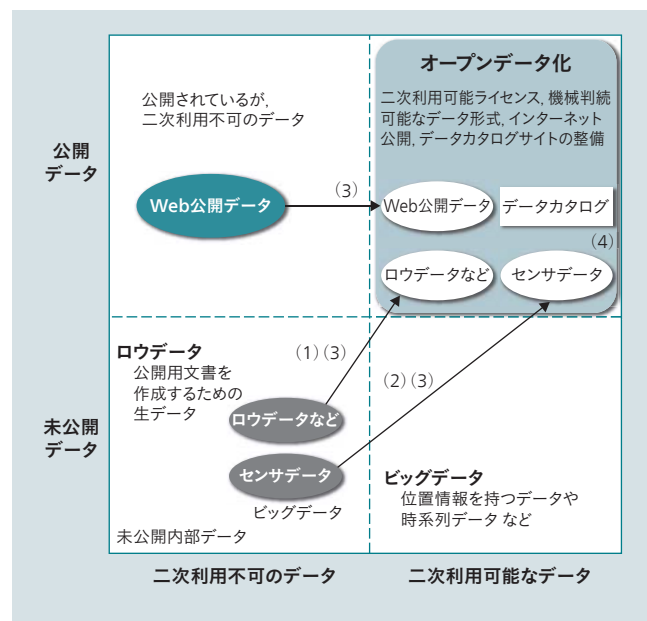


図1 | オープンデータ化の取り組み
公共データのオープンデータ化の考え方を示す。

ウデータ)の公開

(2) センサーなどから観測されたビッグデータ（センサデータ）の公開

(3) 現在のWeb公開データや(1)および(2)の新公開データに対する二次利用可能なライセンス、機械判読可能なデータ形式での公開

(4) どこにどのようなデータがあるかを分かりやすく公開し、データの取得を容易にする「データカタログサイト」の整備

国内外のオープンデータ利活用事例を調査し、分類したものを表1に示す。企業の経営や事業判断への利用、公的機関や地域サービス企業による地域住民サービスへの利用、住民やジャーナリストによる政府や自治体などの活動の評価（透明性）への利用などがある。

表1 | オープンデータの利活用事例の動向

国内外のオープンデータ利活用事例を調査して分類したものを示す。

分類	利用動向	利活用事例
経済の活性化・新事業の創出	企業： 企業の経営や事業判断に利用 • 外部の見える化 • 新市場 • リスク回避	<ul style="list-style-type: none"> • 業界動向や他社動向把握 • 法律文書の改定の把握 • 政府や自治体などのイベント情報のビジネス利用 • 災害・天気・交通・地質等観測・調査などのビジネス利用（リスク分析、市場分析）
官民共同による公共サービスの実現	公的機関や企業： 地域住民サービスや都市設計に利用 • 農業・医療・防災・防犯など	<ul style="list-style-type: none"> • 政府内や自治体間の情報共有 • 災害・天気・交通・地質等観測・調査などの都市設計への活用 • 公共設備ナビの地図サービス • 危険地域、救命救急の地図サービス • バスや電車運行状況の情報提供 • ごみ収集日情報のカレンダー型情報提供 • 地域イベントのカレンダー型情報提供
行政の透明性・信頼性の向上	住民、分析機関、ジャーナリスト： • 政府や自治体などの活動の透明性	<ul style="list-style-type: none"> • 地域の防災速報 • 地域の疾病の広がり速報 • 税金の使途、財務分析

2.2 国内外のオープンデータ推進の動向

英国は、2000年に情報公開法を制定して以来、EU (European Union) 諸国の中でも積極的にオープンデータを推進している。2010年には、公開すべきデータを首相が指定し、省庁ごとにオープンデータ戦略を策定している。米国では、2013年5月にオバマ大統領が、行政機関の情報をオープンデータとして公開することを義務づける大統領令を発令した。また、G8 (主要8か国) の各国政府は2013年6月に「オープンデータ憲章」に合意し、2013年10月までに自国の行動計画を公開することになっている。

日本政府は、2012年12月、内閣官房の高度情報通信ネットワーク社会推進戦略本部 (IT総合戦略本部) に電子行政オープンデータ実務者会議を設置し、行政が保有する地理空間情報、防災・減災情報、調達情報、統計情報などの公共データのオープンデータ化の整備を進めている。経済産業省や総務省なども、関連する会議体の設置や実証事業を推進している。一方、先進的な自治体でも、住民サービス向上や地域活性化を目的に、産学と連携した協議会の設置やイベント開催などを行う動きが見られる。

2.3 オープンデータ流通と課題

オープンデータの流通の概要を図2に示す。オープンデータの公開側には、以下の課題がある。

- (1) 二次利用可能なライセンスとデータ形式の規定
- (2) データカタログ形式の規定
- (3) 行政保有データの効率的なオープンデータ化 (データカタログの作成およびデータコンテンツの構造化)

(1), (2) は内閣官房の関連会議で検討されている。日立グループが開発した (3) に関する対策技術について次に述べる。

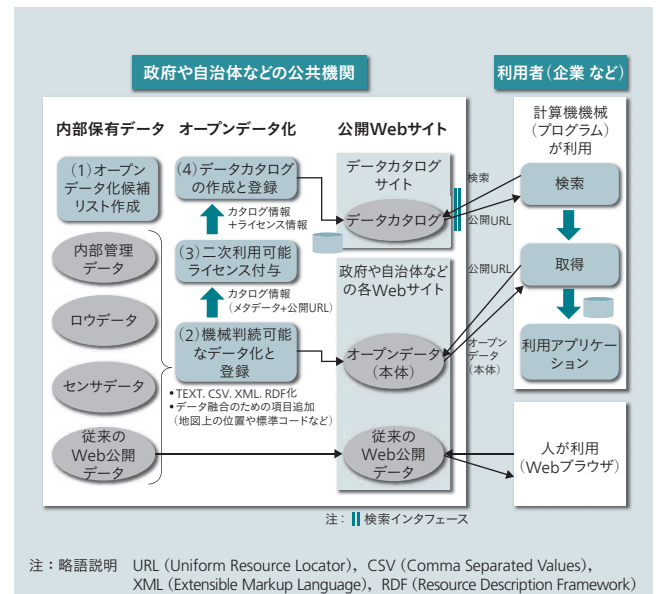


図2 | オープンデータの流通の概要

オープンデータの作成とライセンス処理、登録と公開、検索・取得・利用という上流から下流への流れを示す。

3. オープンデータへの日立グループの取り組み

3.1 オープンデータ化の課題

2013年6月に内閣官房から公開された「電子行政オープンデータ推進のためのロードマップ」での施策の1つに、データカタログサイトの整備がある。データカタログサイトにより、データの分野横断的検索が可能となる。また、データカタログサイトに登録した統計などのデータの内容を機械処理に適した構造のデータ形式で公開し、検索可能にするための整備を推進していくこともロードマップに示されている。これらが実現されれば、例えば、統計データの内容を検索した結果を用いる分析やマッシュアップ(組み合わせ)が容易になる。そのためには、公開するデータコンテンツの構造化が必要となる。

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かか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、オープンデータの構築支援が求められるため、日立グループはこれまで、その支援技術の研究・開発を進めてきた。その中から、データカタログ構築支援技術と、データコンテンツ構造化支援技術について説明する。

3.2 データカタログ構築支援技術

既存のWebサイトで公開されている大量のデータから効率的にデータカタログサイトを構築する技術の研究・開発し、行政機関職員の作業コスト軽減を実現した。

データカタログサイトの構築には、主に2つの作業を要する。1つは、データ自体と、それに関するデータであるメタデータ[データの名前、URL (Uniform Resource

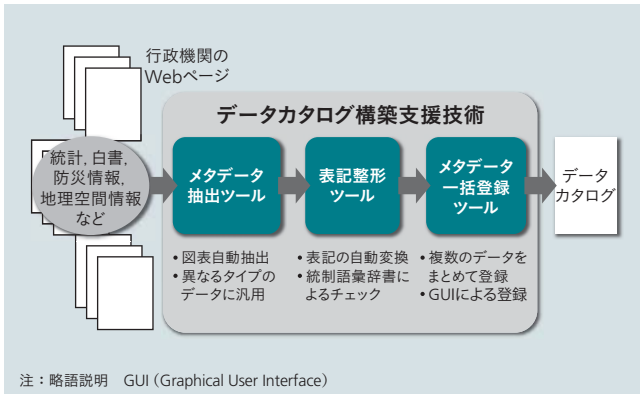


図3 | データカタログ構築支援技術

行政機関が公開するWebページからのメタデータの抽出、抽出したデータの表記整形、整形したデータのデータカタログサイトへの一括登録を自動化できる。

Locator)、データ形式など]の抽出作業である。もう1つは、抽出したデータをまとめてサイトに登録する作業である。しかし、すでに公開されているデータはカタログ化を前提としていないため、メタデータの記載箇所や表現が統一されていない(例えば「平成24年」と「2012年」と「2012」)。そのため、メタデータの抽出や統制に多大な人的コストを要する。また、大量のデータをデータカタログサイトに人手で登録する作業にも多くの時間を要する。このような問題に対し、日立グループは、3つの自動化ツールを開発した(図3参照)。データカタログサイトには、欧米をはじめ各国で構築に使用されているオープンソースソフトウェア「CKAN (Comprehensive Knowledge Archive Network)」を用いた。

「メタデータ抽出ツール」は、Webページに埋め込まれているデータ(図表情報)とそのメタデータを自動抽出することができる。このツールは、さまざまなタイプのWebページに汎用可能であるうえ、抽出結果に対する誤り検出機能を備えているため、人手での誤り修正の手間を最小限に抑えることができる。「表記整形ツール」は、統一表記への自動変換や統制語彙辞書を用いたチェックを行い、表記揺れを防ぐことができる。「メタデータ一括登録ツール」は、CKANのAPI(Application Program Interface)を活用することにより、これまでCKAN既存のGUI(Graphical User Interface)画面で1データずつ行っていた登録作業を、複数データをまとめたファイルを用いて、一括して高速に行うことができる。

この技術を実証するため、政府がWeb上に公開している白書や統計データに適用したところ、メタデータの抽出時間を人手の場合の $\frac{1}{10}$ に、メタデータの登録時間を人手の場合の $\frac{1}{5}$ にそれぞれ低減することができた。

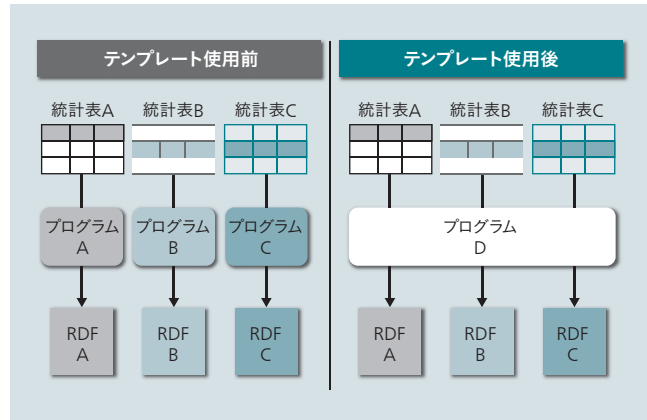


図4 | RDF作成支援ツール

これまで表構造ごとにプログラム作成を要していたが、テンプレートを使用することにより、表構造ごとのプログラム作成が不要となる。

3.3 データコンテンツ構造化支援技術

データコンテンツ構造化支援として、データの内容を機械処理可能な形式へ効率的に書き換えるためのツールを実現した。

このツールは、データをRDF^{※1)}という機械処理に適した構造を持つ形式に変換したものを登録し、検索のためのGUIとAPIを備えたSPARQL Endpoint^{※2)}を提供する。それにより、表やCSV(Comma Separated Values)形式で公開されている統計データの内容への機械アクセス性が向上する。SPARQL Endpointの実現には、事前に統計データをRDFに変換しておく必要がある。統計データは、1つの表であっても数万個の要素から成る場合もあり、RDFへの変換が人手では困難であるため、自動変換プログラムが必須となる。しかし、統計データは、都道府県別の人口、産業コード別の売上高や従業員数など多種多様であり、データに応じてその表構造も多様であるため、表構造ごとに異なるプログラムを作成する必要がある。そのようなプログラム作成を各行政機関の職員が行うには高度なIT(Information Technology)リテラシーが求められるため、日立グループは、さまざまな構造を持つ表をRDFに変換するための汎用的なテンプレートとプログラムを整備し、RDF作成支援ツールを開発した(図4参照)。これにより、表構造ごとに行う必要のあったプログラムの作成コストを低減できる。

この技術を実証するため、政府が公開する統計データに適用したところ、表構造ごとに行う必要のあったプログラムを作成することなく、300万個以上の要素から成るRDFに変換することができ、そのデータを検索できることを確認した^{1), 2)}。

※1) Resource Description Frameworkの略。データの機械処理を目的とした、主語、述語、目的語の3つから成るWorld Wide Web Consortium勧告のデータ形式。
 ※2) RDFのデータを検索するための言語であるSPARQL Query Language for RDF (SPARQL)を使用するためのAPI。

4. オープンデータの展望

4.1 日本におけるオープンデータの方向性

これまで、日本におけるオープンデータ政策の背景や現在の取り組みの状況について述べてきた。

オープンデータの今後の方向性は、電子行政オープンデータ推進のためのロードマップに示されている(図5参照)。ロードマップでは、重要なマイルストーンとして、2015年度末までに他の先進国と同水準のオープンデータの公開と利用を実現することを目標に掲げている。また、ロードマップには、各府省が保有する公共データを横断的に検索可能とするデータカタログサイトの試行版を2013年度に立ち上げ、2014年度以降は本格的に構築・運用していくことが目標とされている。なお、英国や米国をはじめとした世界30か国以上ではすでにデータカタログサイトが開設されており、米国の「Data.gov」では約9.8万件のデータセット(2013年9月時点)が公開されている。2015年に日本が欧米と同水準となるには、2013年6月に閣議決定した「日本再興戦略」³⁾に示されている「2015年度中に世界最高水準の公共データの公開内容(データセット1万件以上)を実現」よりもさらにストレッチした施策の推進が必要となる。

中央官庁だけでなく横浜市や千葉市などの幾つかの自治体では、保有データの公開や産学官による協議会の設置など、先進的な取り組みが見られる。政府のロードマップでは、2013年度中に自治体などの公開データの考え方を整理し、2014年度以降は、有効事例の創出・普及に取り組むことが示されている。

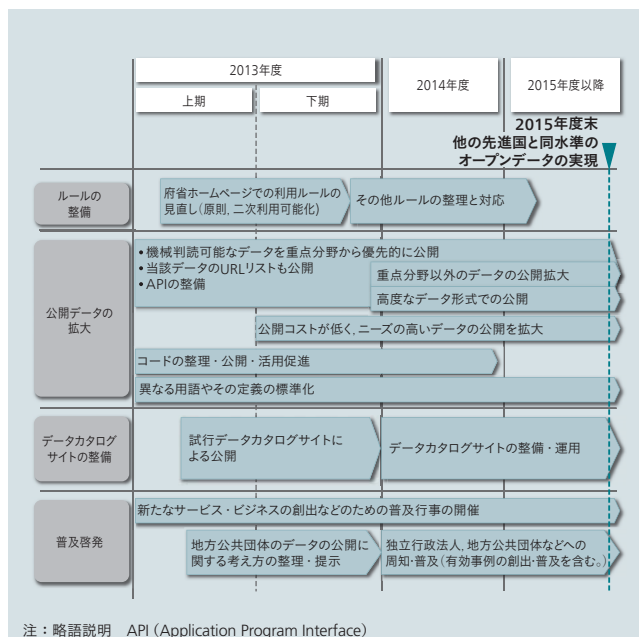


図5 | 電子行政オープンデータ推進のためのロードマップ
各府省による2013年度以降の具体的な取り組みについて内閣官房が作成したロードマップの概要版を示す。

4.2 今後の取り組み

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と、それによる経済活性化などに取り組む動きが加速化している。日立グループは、日本での高水準なオープンデータ環境の実現を支援するために、データ公開側である行政機関やデータ活用側である民間向けに、前述した関連技術やソリューションの整備を進めている。データ公開作業の効率化・省力化のため、メタデータの自動抽出およびデータカタログサイトへの一括登録支援技術、データ活用を容易化するための自動データ収集技術、およびマッシュアップ技術の研究を推進している。日立グループは、オープンデータの進展に応じてこれらの技術の提供することで、顧客の既存業務の効率化や事業サービスの向上支援を図っていく。

5. 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取り組み、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) 武田, 外: 統計データのLOD化とデータ間の関係の表現, 第27回人工知能学会全国大会論文集(2013.6)
- 2) 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(2013.9)
- 3) 日本再興戦略—JAPAN is BACK - (平成25年6月14日) 閣議決定, http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf

執筆者紹介



坂倉 芳崇
2012年日立製作所入社, 情報・通信システム社 公共システム事業部 公共イノベーション事業推進本部 公共ビジネス推進部 所属
現在, 企業情報活用基盤の構築および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
2012年日立製作所入社, 中央研究所 情報システム研究センター 知能システム研究部 所属
現在, オープンデータ実現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



小池 博
1987年日立製作所入社, 株式会社日立コンサルティング 所属
現在, ビッグデータやオープンデータなどを利活用した新ビジネス創生コンサルティングに従事
内閣官房「電子行政オープンデータ実務者会議」構成員(データWG 主査代理)