

オープンデータの活用の現状と課題



中井 亮



眞鍋尚大

CONTENTS

- I オープンデータとは
- II オープンデータ活用の重要性
- III オープンデータ活用の実際：法人データを一例として
- IV オープンデータ活用促進に向けた課題
- V おわりに：日本企業におけるオープンデータの活用に向けて

要 約

- 1 2017年の「世界最先端デジタル国家創造宣言・官民データ活用推進基本計画」の閣議決定以来、日本政府はオープンデータの提供をはじめとするデジタル・ガバメントの実現推進に向けた取り組みを加速させている。特に近年では、官民におけるDXの推進に伴い、オープンデータの重要性はますます高まっている。
- 2 本論考では、海外の先進事例や日本政府におけるオープンデータの取り組みを概観し、DX推進におけるオープンデータの重要性を示したうえで、国内の事例として、経済産業省が運営する法人データベース「Gビジネスフォ」を取り上げ、諸外国の法人データベースとの比較および企業アンケートの結果から、国内における法人データの活用状況・活用意向を整理する。
- 3 また、利用可能性を高める観点から、オープンデータは各組織が保有するデータと自由かつ容易に結合できるよう、LOD（Linked Open Data）の考え方にのっとった形式で提供されることが望ましい。本稿ではLODなどの各種仕様や規格などの技術的背景にも触れ、今後のオープンデータのあり方について、制度・技術の両面から論ずる。

I オープンデータとは

オープンデータとは、原義では、誰でも無料で自由に（二次利用・商用利用も含め）利用することができるデータを意味するものであり^{文献1}、代表的なオープンデータとして、国勢調査の統計データ、国土数値情報、経済センサスなどが挙げられる。

オープンデータは、主に国や地方自治体によって公開されている。日本では、2012年7月に「電子行政オープンデータ戦略」が策定されて以来、各府省庁によりさまざまなデータが積極的に公開されてきた。現在は、デジタル庁が中心となって各府省庁のオープンデータの取りまとめを行っている。

オープンデータは長らく、電子政府の実現による透明性・国民参加の推進といった理念のもとに、中央政府を主体として進められてきた^{文献2}。本稿では主に、オープンデータの活用側、すなわち民間企業や地方自治体の視点から、オープンデータの利活用の可能性やメリット、課題について論じる。

1 世界のオープンデータ推進のきっかけとなった「Data.gov」

オープンデータに通じる概念は古くから存在したが、オープンデータが世界的に注目されるきっかけとなったのは、2009年の米国オバマ政権によって発表された透明性とオープン・ガバメントに対する覚書「Transparency and Open Government」である。同年、米連邦政府は、政府自らが情報公開を進めるため、政府の統計データなどを公開するポータルサイト「Data.gov」を開設した。

オープンデータの取り組みは世界に広が

り、各国でData.govと類似のポータルサイトが公開された。英国では10年1月、オープンデータ公開の5つ星スキームを提唱したティム・バーナーズ＝リー（後述）が推進役となり、「data.gov.uk」を公開した。EUも11年12月、「欧州オープンデータ戦略」を公表し、「政府データを金（gold）に変える」をスローガンとして、オープンデータポータル「PublicData.eu」を開設した。欧米以外では、11年7月にケニアが、アフリカ国家としては初めてのオープンデータポータルとなる「opendata.go.ke」を開設した。また、12年9月には、インド政府が、米政府と共同開発した「OGPL」のプラットフォームを活用して、「data.gov.in」を公開した。そのほかにも、さまざまな国で同様のポータルサイトが公開されている。

こうした流れを受け、13年のG8ロック・アーン・サミットでは、オープンデータに関する議論が行われ、オープンデータ憲章が合意された。

2 「東日本大震災」や「新型コロナ」が契機となった日本のオープンデータ推進

日本では、東日本大震災がオープンデータ推進の機運を高める契機になったといわれる。東日本大震災では、シビックテックコミュニティや企業によるデータボランティアが活躍し、災害情報の周知に貢献した。

震災を通して、オープンデータの重要性が認識される中、日本政府は2012年7月に「電子行政オープンデータ戦略」を策定した。13年1月には、オープンデータポータルサイト「Open DATA METI」のβ版が公開され、同

サイトはその後の「DATA GO.JP」や現在の「e-GOVデータポータル」につながっている。

16年5月、高度情報通信ネットワーク社会推進戦略本部は「オープンデータ2.0」を公表し、オープンデータの基本原則や官民によるオープンデータ推進のための実務者会議設置を定めた。翌17年には、「世界最先端デジタル国家創造宣言・官民データ活用推進基本計画」が閣議決定され、デジタル・ガバメントの実現推進に向けた取り組みがさらに加速した。

20年には、新型コロナウイルス感染拡大をきっかけとして、オープンデータの重要性が再認識された。台湾のマスク配布システムの事例が有名だが、日本でも、東京都が、「新型コロナウイルス感染症対策サイト」を世界に先駆けて開設し、そのソースコードを「GitHub」上で公開したことは注目を集めた。東京都のソースコードは、その後、台湾、米サンフランシスコ市など国内外約50の地域で活用された。

22年6月、デジタル庁によって「デジタル社会の実現に向けた重点計画」が公表されたが、同計画の中で、今後もオープンデータを積極的に推進していくことが示されている。

3 日本のオープンデータの概況

ここでは、わが国におけるオープンデータについて、提供主体別にその特徴を整理するとともに、主要なデータセットを紹介する(表1)。

2012年の「電子行政オープンデータ戦略」の策定、および16年の官民データ活用推進基本法の施行以来、国・地方自治体はオープンデータの公開に積極的に取り組んでいる。府省庁によっては同法施行以前からデータ公開に取り組んでいるところもあり、一例として、国土交通省による「国土数値情報」は遅くとも01年には既に公開されていた^{文献3}。現在はデジタル庁の監督の下、各府省庁が自組織の保有するオープンデータの「棚卸し」を行っている。

地方自治体もまた、オープンデータの利用者であると同時に重要な提供主体である。デジタル庁^{文献4}によれば、すべての都道府県、694の市(全体の87.6%)、527の町(同70.9%)、124の村(同67.8%)が何らかのオープンデータの公開に取り組んでいる。地方自治体によるオープンデータの提供は、その開始時には費用を要するものの、データによる経済効果やイメージアップを考慮すればその投資は短期間で回収できる^{文献5}とされており、

表1 提供主体別に見たデータの特徴

提供者	特徴	実例
国	<ul style="list-style-type: none"> オープンデータ 信頼性が高い 統一されたフォーマット 全国的な分析に向く 	<ul style="list-style-type: none"> G ビズインフォ 国土数値情報 国勢調査
地方自治体	<ul style="list-style-type: none"> オープンデータ フォーマットは自治体によりさまざま データ提供が相対的に不安定 地域に特化した分析に向く 	<ul style="list-style-type: none"> 感染症拡大状況 公共施設・AEDの位置情報 図書館の蔵書情報
民間企業	<ul style="list-style-type: none"> オープンデータの定義からは外れる 申請が必要 大学の研究者のみに利用を限る場合が多い 	<ul style="list-style-type: none"> Yahoo!データセット 楽天データセット クックパッドデータセット

実際に京都市や長野市では情報開示請求件数減少といった効果が現れていることから、今後のさらなる充実が期待される^{文献6}。

一方、民間企業によるオープンデータの公開は十分に進んでいるとは言い難い。国立情報学研究所の情報学研究データリポジトリにはさまざまな民間企業から提供されたデータセットが掲載されているケースも見られる^{文献7}が、多くは提供先を大学などの研究機関に限っており、また二次利用も制限されていることから、厳密にはオープンデータの定義には合致しない。

II オープンデータ活用の重要性

1 オープンデータの利点と欠点

利用者の立場から考えると、オープンデータにはさまざまな利点と欠点がある。代表的な利点は費用である。たとえば資金力の乏しい企業にとって、あるいはプロジェクト立ち上げ直後の予算の少ない時期に分析の試行を行う場合には、無料であることは大きな利点となるだろう。

オープンデータは多くの場合、国・地方自治体から提供されており、信頼性が高いことも利点である。特に、公共政策にかかわる分野や採算の見込めない分野では、オープンデータが唯一の情報源となる場合が多い。一方で、民間企業からも有償でデータが提供されている分野（一例として交通・人流データなど）においては、有償データと比較すると、サポートや更新頻度、精度・粒度の点では劣る場合もある。

また、データそのものは誰でも利用可能であるため、データ単体で価値を創出すること

は、必ずしも容易ではない。自社が独自に保有するデータと組み合わせたり、高度な分析手法やユニークな活用方法を導入したりするなどの工夫が求められる^{文献8}。

2 クローズドデータの限界、

オープンデータとの併用の可能性

オープンデータに対して、各企業などの組織内のみにおいて蓄積・管理・利用されるデータをクローズドデータと呼ぶ。典型的な例としては、取引先や顧客の一覧、あるいは取引・財務情報などが挙げられる。

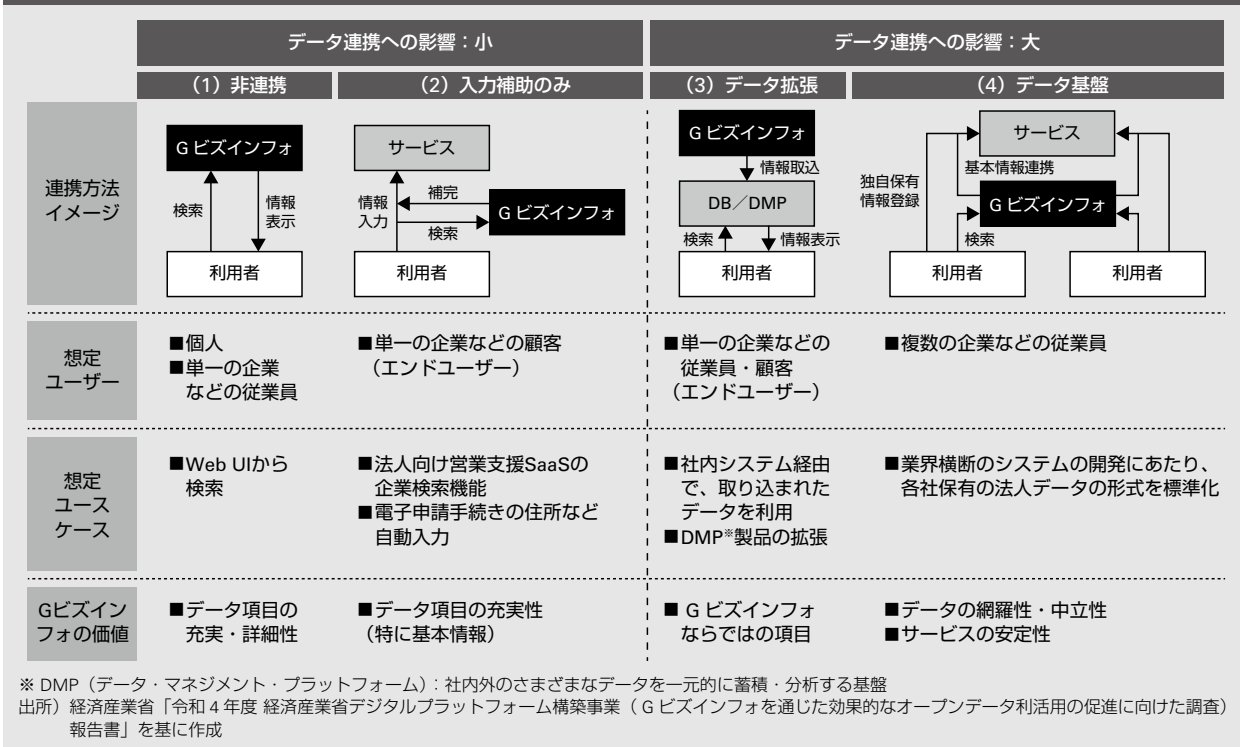
既に多くの企業ではクローズドデータの分析が行われているが、クローズドデータのみを利用してデータ活用を推進することには限界がある。たとえば、新規に取引先や顧客を開拓したい場合、既に保有しているデータを利用するだけではアプローチの範囲に限られる。

このような場合は、自社の保有するクローズドデータを活用して既存のビジネスの傾向を分析しつつ、そのようにして得られた知見を外部の情報源から得たデータと組み合わせることで、それまで対象となっていなかった領域へビジネスを拡大することが可能になると考えられる。こうしたデータ活用の際に必要な外部情報源の一つとして、オープンデータは有効な選択肢となり得る^{文献9, 10}。

3 オープンデータのユースケース

ここでは、主に民間企業におけるオープンデータの活用方法について、いくつかの事例を紹介しながら、分類を行う。以下では、「オープンデータと自社のデータをどのように組み合わせるか」という基準で、ユースケースを4段階に分類する（図1）。

図1 オープンデータのユースケースの4分類



(1) 「非連携」パターン

ユースケースの1段階目は「非連携」パターンである。オープンデータを提供元のサイトから直接利用したり、あるいはオープンデータを単独で（場合により加工を加えて）提供したりする場合が該当する。典型的な例としては、G Bizインフォ（gBizINFO）やe-StatなどのWebサイトで直接検索してデータを閲覧またはダウンロードするケースがある。また、自治体のオープンデータの活用事例としてよく見られる、地域の特徴的な地物・観光地などを地図上に表示するサービスも、このケースに該当する。

(2) 「入力補助のみ」パターン

UI（ユーザー・インターフェイス）上で簡易的なオープンデータの検索を行い、検索

結果を利用するパターンである。たとえば、システムの利用者が住所・法人名などを入力する際に、直接入力するのではなくG Bizインフォの検索結果を利用して入力の手間を省くといった実装が考えられる。卑近な例では、通信販売や会員登録のWebサイト上で、郵便番号を入力すれば住所が補完される機能をイメージすると分かりやすい。このパターンでは、オープンデータは利便性に寄与しているものの、オープンデータそのものが自社の保有データと連携して利用されているわけではない。

(3) 「データ連携」パターン

自社で保有するデータベースにオープンデータを組み込むなどの方法で自社データとオープンデータを紐づけて利用する場合は、こ

のパターンに該当する。たとえば、取引先リストと法人オープンデータを結合することにより、取引先の企業情報（商号、本店所在地、従業員数など）を定期的に自動更新するユースケースや、自社で保有している各支店の売上情報を地理オープンデータと結合することにより、地域ごとの業績を分析したり、業績情報と距離や人口などの情報を組み合わせて支店の配置を最適化したりするユースケースが該当する。

(4) 「データ基盤」パターン

前掲の「データ連携」パターンの対象を複数社に拡大したものであり、オープンデータをデータの「基盤」もしくは「仲介者」として活用するパターンである。このパターンは、業界横断での情報処理システムや、同業種の複数企業を顧客とするSaaSなどを構築する際に有用である。このような場合には、各企業が保有するデータの体系を共通化することが求められるが、その際、特定の企業に依存せずに各企業のデータを仲介するデータ体系として、オープンデータを利用することができる。

4 オープンデータの価値

ここでは、データ連携への影響度によって4つのパターンを整理したが、オープンデータに求められる価値（利用者視点で言い換えるなら、オープンデータを利用すべき理由）は、パターンによって異なる。パターン1、2では、データが無料であること、加工しやすいこと、システム連携のためのAPI（アプリケーション・プログラミング・インターフェース）を利用しやすいことなどが挙げられ

る。パターン3、4ではこれらの点に加え、データ項目の正確性や充実度、さらにはデータが一企業に依存しない中立的なものであることも価値と見なされるであろう。

Ⅲ オープンデータ活用の実際： 法人データを一例として

1 法人データベースのビジネス活用

これまで、オープンデータ全般について語ってきたが、オープンデータ活用について、より具体的な説明をするため、本章では法人データに絞って説明したい。これから説明する内容は、2022年度に野村総合研究所（NRI）が経済産業省大臣官房デジタル・トランスフォーメーション室とともに実施した「令和4年度経済産業省デジタルプラットフォーム構築事業（Gビジネスフォを通じた効果的なオープンデータ利活用の促進に向けた調査）」が基になっている。

ここでいう法人データとは、法人名、法人番号、所在地などの法人基本情報、売上、利益、資産などの財務情報、その他、職場情報や特許の取得状況、過去の表彰、信用スコアなど法人にかかわる情報全般を指す。

企業や公的機関は、事業活動を通して、顧客あるいはパートナーとなる法人のデータを収集・蓄積しているが、外部の法人データベースを活用することで、さまざまな業務を効率化、高度化できる。

たとえば、法人データは、信用調査に活用されている。多くの企業では、信用調査会社が提供する信用スコアや過去の活動状況などの法人データを基に、取引してよい企業かどうかを判定している。また、新規事業を立ち

上げる際の市場分析として、法人データベースから、ターゲット業界に属する企業の財務データなどを収集・分析することで、市場規模を推定している。さらに、有力な競合企業を明らかにする際にも、法人データベースを活用することが多い。

法人データベースはデータの閲覧・ダウンロード以外にも、APIを提供しているため、自動でデータを取得し、自社データベースに結合することができる。そのため、自社が持つ法人データと組み合わせて、既存顧客の詳細を把握したり、新規顧客のターゲットリストを作成したりすることもできる。また、自社データベースでは、「表記揺れ」などの問題で正確なデータが入力されていない箇所について、法人データベースの情報から「名寄

せ」することもよく行われている。

2 民間データベースと政府系データベース

法人データベースは大きく分けて、信用調査会社や経済メディア、マーケティング企業などが提供する民間データベースと、政府組織が提供する政府系データベース（オープンデータベース）の2種類が存在する（表2）。

民間データベースとしては、帝国データバンクが提供する「COSMOS」、東京商工リサーチが提供する「Data Approach」、ユーザベースが提供する「SPEEDA」などが有名である。

帝国データバンクや東京商工リサーチなどの信用調査会社が提供する法人データベース

表2 国内の主要な法人データベース^{注1}

データベース種類	民間データベース	民間データベース	民間データベース	オープンデータベース
データベース名称	COSMOS	Data Approach	SPEEDA	G ビズインフォ
運営主体	帝国データバンク	東京商工リサーチ	ユーザベース	経済産業省
掲載企業	国内・海外を含めて約4.3億社	国内・海外を含めて約1.8億社	約1,000万社（うち国内が約150万社）	国内全法人
データ項目	<ul style="list-style-type: none"> 企業コード 商号 所在地 産業分類 仕入先／販売先 評点 など 	<ul style="list-style-type: none"> 会社概況 評点 業種分類 担当者情報 主要取引先 財務諸表 など 	<ul style="list-style-type: none"> 会社概要 財務情報 市場データ 開示資料 IRデータ ニュース／人事情報 など 	<ul style="list-style-type: none"> 会社概要 財務情報 補助金情報 調達情報 特許情報 くるみん えるぼし認定 など
データ形式	<ul style="list-style-type: none"> カード、リスト、アドレスブック、CD-R（オンライン提供あり） APIによるデータ提供あり 	<ul style="list-style-type: none"> Excel／Access／CSV／TXTなどの形式でデータ提供 APIによるデータ提供あり 	<ul style="list-style-type: none"> Excel形式でエクスポート可能 	<ul style="list-style-type: none"> CSV形式でエクスポート可能 APIによるデータ提供あり（REST APIとSPARQL API）
データソース	<ul style="list-style-type: none"> 各種独自調査や有価証券報告書などの公開情報 ビューロー・ヴァン・ダイクと提携 	<ul style="list-style-type: none"> 各種独自調査や有価証券報告書などの公開情報 D&B社と提携 	<ul style="list-style-type: none"> EDINET、TDNET 東京商工リサーチやDow Jones社などと提携 	<ul style="list-style-type: none"> 国税庁 法人番号サイト EDINET、官報 総務省 統一資格審査サイト 厚労省しよくばらぼ
独自の強み	<ul style="list-style-type: none"> 信用調査会社として日本最大で、調査によって常に最新の情報に更新 	<ul style="list-style-type: none"> 「評点」など独自指標を提供 D&B社と提携し、世界標準のフォーマットで情報を提供 	<ul style="list-style-type: none"> 複雑な画面を廃し、シンプルな操作性を追求 業種データを利用して、競合分析を自動化 	<ul style="list-style-type: none"> 政府系データベースならではの多様な情報がそろっている 無料で利用可能

は、信用調査を通して入手した豊富なデータ（非上場企業を含む）がそろっていることが特徴である。一方で、ユーザーベースが提供する「SPEEDA」は、公開情報が中心だが、市場分析などのユースケースを意識した利用のしやすさに定評がある。

一方、政府系データベース（オープンデータベース）としては、経済産業省が運営する「G ビズインフォ」、金融庁が運営する「EDINET」などがある。

「G ビズインフォ」は、さまざまな府省庁が保有する法人データを集約したデータベースである。法人ごとに、法人基本情報、財務情報、職場情報などがまとまっている。「EDINET」は、有価証券報告書などの開示書類に関する電子開示システムのことを指すが、APIを利用することでデータ取得にも活用できる。

そのほか、法人データを取得できるデータベースとしては、デジタル庁が運営する「e-Govデータポータル」が挙げられる。こちらは、法人データに限らず、公的機関が保有するデータを集約したデータカタログサイトであり、「中小企業実態基本調査」「外資系企業動向調査」などの法人にかかわるデータを取得することが可能である。

民間データベースは、取得できるデータ項目が豊富であり、ユースケースごとに使いやすいUIが整備されているが、利用料金が高額である。一方で、オープンデータベースは、データ項目やUIでは見劣りするが、無料で使用でき、補助金の取得情報、政府事業の調達状況、ESG関連情報（くるみん認定、えるぼし認定）など政府系ならではの情報が充実していることが特徴である。また、登記情報が基になっているため、すべての法人を

網羅しているという特徴もある。両者とも一長一短があり、利用目的や利用者の状況によって、使い分けたい。

3 オープンデータベースの課題： 「G ビズインフォ」を例にして

ここまで、政府系データベース（オープンデータベース）と民間データベースの特徴について述べたが、以下では政府系データベースの課題について、経済産業省が提供する法人オープンデータベース「G ビズインフォ」を例に挙げて論ずる。

NRIでは、2023年1～2月に、前述の調査事業の一環として、G ビズインフォの利用者および一般の法人を対象に、法人データベースに関するアンケート調査を行った（図2）。

アンケート対象は、地方自治体や信用情報調査会社など、一般的な企業よりもデータへの関心が高いと思われる業種を中心としたが、それにもかかわらずG ビズインフォの認知度は約25%であり、現に利用しているユーザーの割合は約4%にとどまった。

業種別の利用率では、「信用情報・法人データベース提供会社」を除けば、当初G ビズインフォを活用していると考えられていた諸業種を除く「一般の民間企業」が最も高かったが、法人データを専門に扱う企業は既に民間の企業データベースを利用しているため、現状ではG ビズインフォを利用する利点を感じていないことが理由と考えられる。

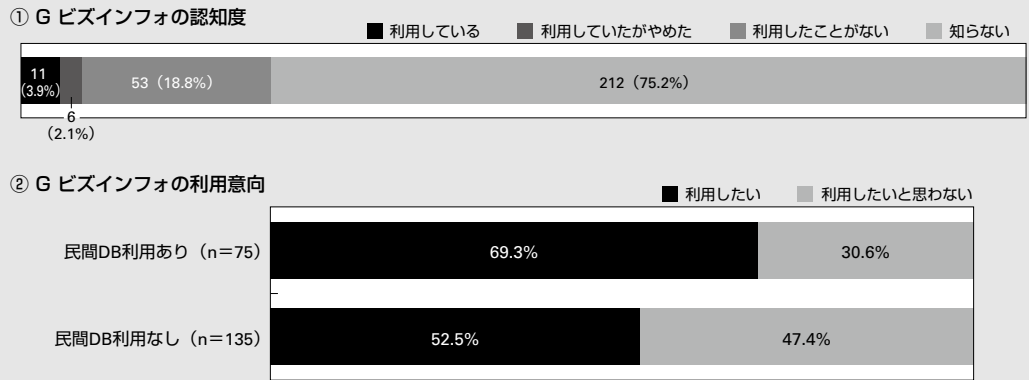
一方で、未利用者におけるG ビズインフォの利用意向は約6割と高かったことから、認知度の向上が今後の課題になると考えられる。前節で述べた政府系データベースならではの特色に関連して、G ビズインフォ特有の

データ項目を充実させることができれば、企業に対するアピールポイントとなるであろう。

G ビズインフォの利用意向に関しては、既に民間の企業データベースを利用している

企業の方が、利用していない企業よりも約2割高かった。この点から、G ビズインフォ独自の機能あるいはデータが求められていることが示唆される。また、認知・利用の促進

図2 G ビズインフォにかかるアンケート調査結果



出所) 経済産業省「G ビズインフォを通じた効果的なオープンデータ利活用の促進に向けた調査」(2022年度)より作成
 ※ 同調査の一環で実施したアンケートを基に集計
 アンケート対象: 地方自治体、大学、法人向けSaaSベンダー、データ活用推進企業(調査会社など)、信用情報・法人データベース提供会社、金融機関などの民間企業
 実施時期: 2023年1月20日～2月24日
 有効回答数: 282件

表3 G ビズインフォに対する利用者の評価

		あるべき姿	実態	利用者からの評価
データ	完全性	全法人について、主要なデータ項目が漏れなく利用可能	登記された法人すべてを収録しているが、データ項目には抜け漏れもある	○ 民間データベースと異なり、全法人のデータが網羅されていることへの評価は高い △ 個別の法人について、項目によっては情報が欠けている点は、利用者から改善要望があった
	信憑性	各府省庁が保有するデータの一括提供を行うことで、デジタル・ガバメントの基盤を構成する	政府提供データと、各法人が入力したデータが混在。出所記載もあるが、慣れていない利用者には判別が難しい	○ 政府提供であることへの安心感・評価は高く、利用するきっかけとなっている △ 一部には、政府提供ではなく各法人が自己入力しているデータも存在しており、自己入力データの信憑性に懸念を抱く利用者の声も聞かれた
	最新性	できるだけ新しいデータを提供する	連携タイミングの関係上、更新にタイムラグが生じ得る	△ タイミングによっては、古い情報が更新されないまま残っており、ほかのデータベースとの矛盾が生じている、と指摘する利用者もいた
	一貫性	表記が定まっており、データの出典によらず形式が統一されている	情報収集元が多数あるため、ある程度の表記揺れなどが避けられない	△ 自社システムに組み込み、業務を効率化する目的から、代表者名などの表記揺れの解消が望まれる
機能	REST API	APIの開放により、開発者によるデータ利用を促進している	一部でAPIの安定性に対する不安がある	○ 他府省庁が提供するAPIよりも安定性・利便性が高いため、G ビズインフォAPIに移した、という開発者の声があった △ 一方で、G ビズインフォAPIについても、時折サーバー側の問題によるエラーが返される、という指摘もある
	SPARQL API	SPARQL APIの開放によりLinked Open Dataとしての利用を促進する	法人データにおいてはLODの普及度が低いいため、活用されているとはいえない	△ インタビュー先のうち多数は「知らない」・「知っているが使いにくいので使わない」といった評価であった

出所) 経済産業省「G ビズインフォを通じた効果的なオープンデータ利活用の促進に向けた調査」(2022年度)より作成

に当たっては、オープンデータの利活用に関する啓発活動やマニュアルの整備といった取り組みも有効と考えられる。

本調査事業では、前述のアンケート調査に加え、現在G Bizインフォを活用している利用者を対象としたヒアリング調査を行い、その結果をデータおよび機能の2つの観点から取りまとめた(表3)。データの完全性、信憑性などについて、高く評価する声があったが、より高い信憑性を担保するためには、政府提供データと自己入力データの区別、あるいは自己入力データに対する認証機能の付加などの検討が必要となるだろう。

一方で課題も見られ、特に、一貫性においては、代表者名や業種などの情報に表記揺れが多いため、自社保有データとの紐づけが困難であるという意見があった。利活用の推進に当たっては、ほかのデータベースとのデータ連携を前提とした、組み合わせやすい形式でのデータ整備が求められている。

IV オープンデータ活用促進に向けた課題

前章ではG Bizインフォを例に、オープンデータベースの課題について言及した。ここでは、それらの課題のうち、①オープンデータの普及促進、②オープンデータ活用を支える技術導入：適切な形式での公開による利活用の促進、について、先進事例や先端技術を紹介しながら、課題解決策を提案したい。

1 オープンデータの普及促進：諸外国における取り組み事例

前章の最後では、日本の政府系データベ

ス(オープンデータベース)にはいまだ認知度向上の余地があり、普及に向けた活動が重要であることを述べた。ここでは、オープンデータの普及に関する参考として、欧米のいくつかの先進的な取り組みに触れる。

欧州連合出版局が運営するサイト「data.europa.eu」は、EU加盟27カ国に関する150万件以上のデータセットをさまざまな形式で提供しているだけでなく、オープンデータの基礎知識をPDF形式の文書やYouTube動画教材として公開している。加えて、「EU Datathon」と題してオープンデータを活用したアプリ開発のコンテストを行うなど、普及活動にも積極的に取り組んでいる。

また、米国ニューヨーク市のオープンデータポータルは、「郵便番号区域別の電力消費量」や「地区別の廃棄物収集量」など、市ならではの詳細な単位での統計データを提供することに加え、こちらもアプリ開発コンテストを実施しており、優秀作品の制作者には市から賞金が贈呈されるだけでなく、コンテストでアプリが注目を得ることによりベンチャーキャピタルから投資を受けた例もあるという。

日本においても諸外国の事例を参考として、オープンデータの活用に関する教育・啓発、および、活用を推進するためのインセンティブ設計が求められる。

2 オープンデータ活用を支える技術導入

認知度の向上に加え、もう一つの重要な点が、適切な形式での公開による利活用の促進である。オープンデータは、ただ公開すればよいというものではない。無論、どんな形式

表4 オープンデータの5段階

段階	要件	形式の例
★	どんなフォーマットであれ、Web上で利用可能であること。ただし、誰でも利用できる自由なオープンライセンスのもとで提供されていること	PDF、HTML、紙をスキャンした画像データなど
★★	コンピュータが判読可能な形式であること	Excelなど
★★★	上記に加えて、特定の商用ソフトウェアに依存しない形式であること	CSVなど
★★★★	上記に加えて、W3C ^{※1} が規定したRDFやSPARQLといった規格に基づいて、URI ^{※2} で情報を指定すること	RDF、XMLなど
★★★★★	上記に加えて、データの文脈を明らかにするために、ほかのデータへリンクすること	Linked Open Dataとしての要件を満たすRDF、XMLなど

※1) W3C : World Wide Web Consortium. Web技術に関する標準化団体
 ※2) URI : Uniform Resource Identifier 情報源などのリソースを一意に表すための文字列。
 出所) Berners-Lee, Tim. 2006/2009. "Linked Data" <https://www.w3.org/DesignIssues/LinkedData.html> (2023年6月24日閲覧) を参考に作成

であれ、公開しないよりは公開した方がよいが、利活用を促進するためには、よりコンピュータで扱いやすく、自由に利用可能で、曖昧性が排除され、ほかのデータと組み合わせやすい形式で提供することが望ましい。

提供形式を検討する上では、WWW（ワールド・ワイド・ウェブ）の考案者でもあるティム・バーナーズ＝リーにより提唱された「5つ星オープンデータ」（表4）のスキームが指針となる。

3 オントロジーによる表現と Linked Open Data

オープンデータの3つ星までは、データ分

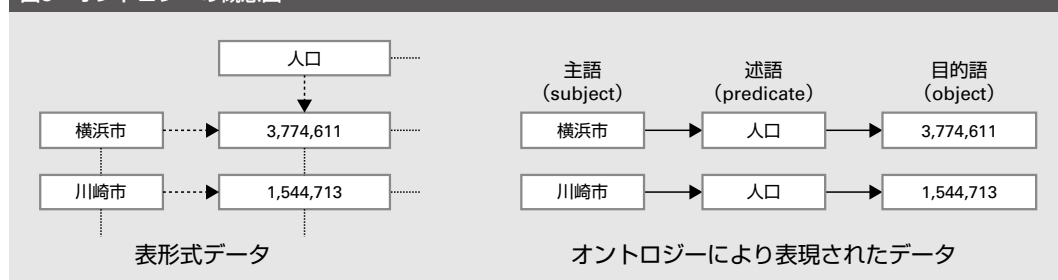
析に携わったことがある読者ならば経験上、感覚的に理解しやすいであろう。ここからは、4つ星以上のデータを扱ううえで重要となる概念、「オントロジー」と「LOD (Linked Open Data)」について解説する。

オントロジーとは、「主語・述語・目的語」の3組によるデータの表現方法であり、ファイル形式としては、RDF (Resource Description Framework : メタデータを記述する枠組み) の仕様にのっとりXML (Extensible Markup Language : 任意のデータ構造に対応した記述言語) ファイルにより提供されることが多い文献¹¹。

多くの読者には、データの表現方法として、表形式の方がなじみ深いであろう。オントロジーによる表現と表形式との差異を図3に示す。この例では、オントロジーの特徴が分かりにくいかもしれないが、オントロジーの利点の一つは、図4のような1対多対応を表現できることである¹²。

ここまでが4つ星に対応するデータ表現形式であるが、5つ星のデータにはLODとして、ほかの情報源とリンクすることが求められる。LODの考え方・メリットを図5に示す。それぞれの主語や述語に対して、文字列による表現だけでなく外部情報源へのURI (表4※2参照) をリンクとして付与することにより、たとえ情報源によって文字列表現が異なっても同一の概念を示すことがで

図3 オントロジーの概念図^{注3}



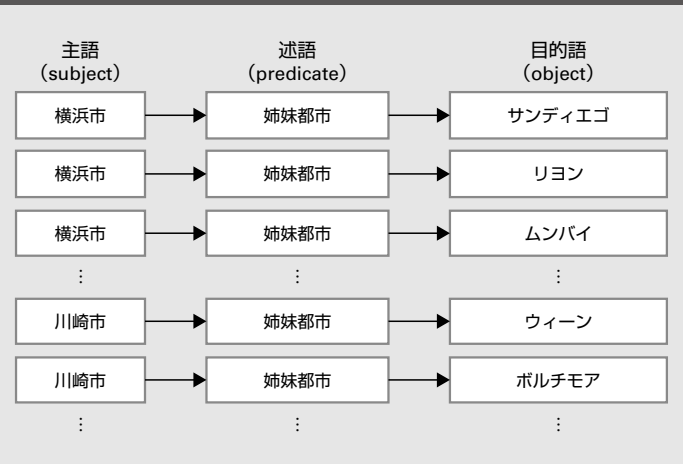
きるため、さまざまな情報源から得た情報を効率的に組み合わせて利用することが可能となる。

4 Linked Open Dataの活用

前述のように、主語・述語がリンク可能な形で定義されたデータにはさまざまな活用可能性がある。たとえば、SPARQLのような問い合わせ言語を用いて、「神奈川県に所在する自治体の首長の一覧を出力せよ」「神奈川県に所在するそれぞれの市の友好都市のうち、人口が100万人以上である都市の名前を出力せよ」といった内容を表す問い合わせ（クエリ）を組み立て、その結果を機械判読可能な形で利用することができる。

もっとも、生成系AIが実用化されつつある現代では、このような複雑な問い合わせができる利点は薄れつつあると思われるかもし

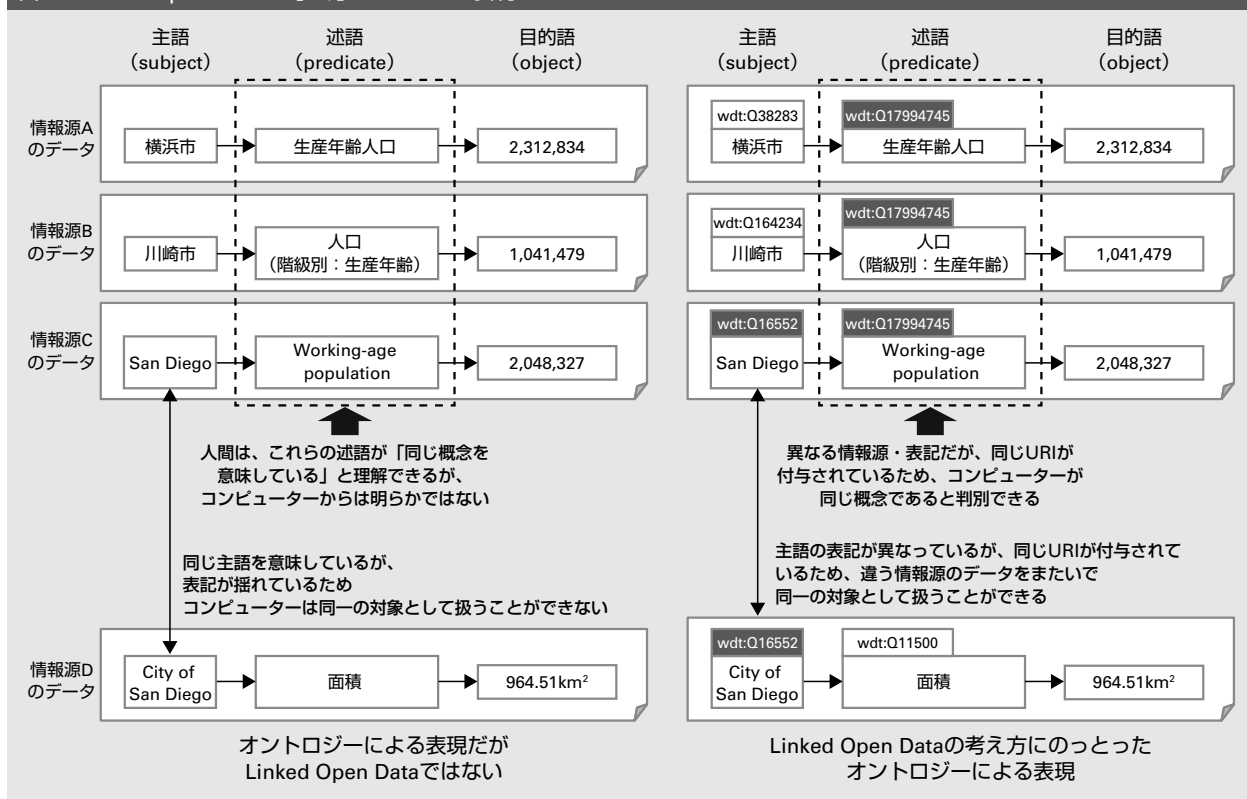
図4 オントロジーによる1対多対応の表現



れない。しかしながら、AIによる誤りや質問文の解釈の曖昧性を排除したうえで、かつ、ほかの情報源と組み合わせ可能な形でデータを扱うことができる点は、依然としてLOD特有の利点であるといえるだろう。

企業活動において、より実用的な例として

図5 Linked Open Dataの考え方にのっった表現³⁴



は、人的あるいは企業間ネットワークの表現が考えられる。株主・役員の情報 LODとして表現すれば、「ある企業 X の役員 A 氏は、別の企業 Y の役員でもある」「企業 Y の役員には B 氏もおり、B 氏は企業 Z の大株主である」といった情報をコンピュータで扱い、容易に検索することができる。このような情報は、紹介を通じた新規営業・取引先の開拓や、コンプライアンスチェックにも活用することができるであろう。

また、オープンデータとクローズドデータを組み合わせた Linked Data の活用事例としては、企業内において特許・論文・技術キーワード・プレスリリースなどの公開情報に加え、人物名・顧客導入事例といった内部情報を Linked Data として整理し、検索可能なナレッジマネジメントシステムを構築した事例がある^{文献12}。

V おわりに：日本企業におけるオープンデータの活用に向けて

本稿では、日本を中心とした世界のオープンデータの現状について、いくつかの具体例を挙げながら論じてきた。オープンデータはそれ自体が企業活動における価値の源泉となるだけでなく、それと組み合わせることにより、自社で保有するクローズドデータにも新たな価値を付加できる。

企業活動における自社保有データ活用の有用性は広く認識されているところであるが、一方で、オープンデータそのものの存在や、オープンデータをクローズドデータと組み合わせることの利点は、十分に認知されているとは言い難い。オープンデータをまだ活用して

いない企業は、自社業務における活用の可能性を検討したうえで、第 II 章で整理したユースケースの段階に沿って、まずは自社の業務の補助として使い、徐々に自社のデータを管理するシステムへの組み込みやデータプラットフォームの構築など、高度な利用へと拡大することが望ましい。その際、活用のレベルに対応した技術を選定することも重要となる。

企業がオープンデータを活用するためには、適切なデータ提供が不可欠である。政府には今後も、オープンデータの積極的な推進が求められることはいうまでもない。地方自治体も、デジタル庁が提供する「オープンデータパッケージ」などの資源を活用しながら、可能な範囲で提供していくことが望まれる。また、データの提供にとどまらず、活用事例の紹介やマニュアルの整備、インセンティブの付与など、利用促進のための取り組みも重要である。技術面では、LODをはじめとする規格にのっとった形式で提供することにより、より高度な利活用を促進できる。

さらに高度な段階としては、企業がオープンデータの利用者として活用するのみならず、提供者となってデータプラットフォームを構築することも考えられる。今後、生成系 AI をはじめとするさまざまな技術の進歩を背景に、データ利活用がますます重要になるにしたがって、企業活動におけるオープンデータへの取り組みは大きなアドバンテージとなるであろう。

注

- 1 表中に挙げた以外にも、ユーソナーの法人データベース「LBC」は有名である。
- 2 このようなデータを表形式で表現しようとすれ

ば、姉妹都市の数だけ「姉妹都市1」「姉妹都市2」……といった列をつくらなければならないだろう。さらに、もし姉妹都市が増えれば表の列を増やさなければならず、その結果としてデータの定義をつくり直す必要に迫られるかもしれない。オントロジーによる表現ではそのような必要は生じない。

- 3 図中のデータは各種公開情報に基づくが、最新のデータとは限らない。以下同じ。
- 4 wdt:Q~に続く数字はWikidata (<https://www.wikidata.org/>) におけるURLの短縮形を表す。

参考文献

1. 大向一輝. 2013. 「オープンデータとLinked Open Data」『情報処理』54 (12): 1204-1210, 情報処理学会
2. 本田正美. 2014. 「『電子政府』の変遷と到達点としてのオープンガバメント・オープンデータ」『情報システムと社会環境研究報告』2014 (3): 1-6, 情報処理学会
3. 国土交通省, 2001. 「国土数値情報ダウンロードサービス」Internet Archive による<https://nlftp.mlit.go.jp/ksj/> の2001年6月3日のアーカイブ <https://web.archive.org/web/20010603121037/https://nlftp.mlit.go.jp/ksj/> (2023年6月24日閲覧)
4. デジタル庁. 2023. 「オープンデータ取組自治体一覧 (令和5年3月1日時点)」https://www.digital.go.jp/resources/data_local_governments/ (2023年6月24日閲覧)
5. 吉田暁生, 野田哲夫, 本田正美. 2016. 「地方自治体におけるオープンデータの活用の効果と課題」『山陰研究』9: 97-109, 島根大学法文学部山陰研究センター
6. デジタル庁. 2023. 「オープンデータ100」https://www.digital.go.jp/resources/data_case_study/ (2023年7月21日閲覧)
7. 国立情報学研究所. 2023. 「情報学研究データリポジトリ」<https://www.nii.ac.jp/dsc/idr/datal-ist.html> (2023年6月24日閲覧)
8. 東富彦. 2014. 「オープンデータビジネスの事例と分析」『智場』119: 28-39, 国際大学グローバルコミュニケーションセンター
9. Kaasenbrood, Maaïke; Zuiderwijk, Anneke; Janssen, Marijn; de Jong, Martin; Bharosa, Nitesh. 2015. "Exploring the Factors Influencing the Adoption of Open Government Data by Private Organisations". *International Journal of Public Administration in the Digital Age* 2 (2): 75-92.
10. Gottfried, Anne; Hartmann, Caroline; Yates, Donald. 2021. "Mining Open Government Data for Business Intelligence Using Data Visualization: A Two-Industry Case Study". *Journal of Theoretical and Applied Electronic Commerce Research* 16 (4): 1042-1065.
11. Ejigu, Dejene; Scuturici, Marian; Brunie, Lionel. 2007. "An Ontology-Based Approach to Context Modeling and Reasoning in Pervasive Computing". In *Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops* 14-19.
12. 井形伸之, 他. 2013. 「Linked Dataを用いた情報統合・活用技術」『Fujitsu』64 (5): 464-470.

著者

中井 亮 (なかいりょう)

野村総合研究所 (NRI) CXコンサルティング部シニアコンサルタント

専門は機械学習 (自然言語処理、時系列データ処理) のビジネス応用、プライバシー/データ政策、図書館情報学

眞鍋尚大 (まなべなおひろ)

野村総合研究所 (NRI) CXコンサルティング部シニアコンサルタント

専門は、社外データベース活用、CX高度化支援、機械学習・統計モデリング