

オープンデータの高度利用を支える Linked Open Dataとその応用

Linked Open Data for Enhanced Use of Open Data and New Applications

● 井形伸之 ● 桑 照宣 ● 中澤克仁 ● 塩田哲義

あらまし

「世界最先端IT国家創造宣言(2013年6月閣議決定)」では、公共データの民間開放(オープンデータ)を推進するため、2014年度および2015年度の2年間を集中取組み期間と位置づけており、官公庁・自治体を含めたオープンデータ活動が広がっている。こうした中、富士通研究所では、オープンデータの利用価値を高める技術として、オープンデータの五つ星と言われるLinked Open Data(LOD)に関する技術開発を早くから開始し、研究成果の一部を一般公開している。

本稿では、Web上のLODを検索するサービス(LOD4ALL)や地域の特性を見える化するWebツール(EvaCva)を中心に、オープンデータのエコシステム創出に向けた富士通研究所の取り組みを紹介する。

Abstract

In Japan, the Declaration to be the World's Most Advanced IT Nation (approved by the Cabinet in June 2014) designates the two-year period between 2014 and 2015 to be an intensive development phase in promoting the use of public data (open data) in the private sector. In line with this, an increasing number of bodies including governmental offices and local authorities are rolling out initiatives to leverage open data. Fujitsu Laboratories has been engaging in the development of technology to enhance the value of open data for some time. It focuses on the utilization of linked open data (LOD), which is recognized as 5-star open data, and their research results are partly made accessible to the general public. This paper introduces some of Fujitsu Laboratories' projects for the creation of an open data eco-system, including a Web-based LOD query service (LOD4ALL) and a Web tool to visualize region-specific information (EvaCva).

まえがき

近年、目的を問わず誰もが無償で二次利用できる「オープンデータ」が注目されている。日本でも「電子行政オープンデータ戦略」が2012年6月に策定されて以降、公共データの民間開放を推進する動きが活発化しており、官公庁・自治体を中心にオープンデータ活動が広がっている。2014年10月には、政府のオープンデータリストである「データカタログサイト」⁽¹⁾が正式運用され、2015年3月現在では100を超える自治体がオープンデータ活動に取り組んでいる。

オープンデータ活動は世界中に急速に広がっており、英国の非営利団体Open Knowledge Foundationでは、世界各国のオープンデータの進捗状況をいくつかの評価尺度から点数付けし、進捗度ランキングを公開している。⁽²⁾このランキングによれば、英国を筆頭にEU諸国がトップランナーであるが、日本も2013年の27位から2014年12月には19位へと順位を上げている。

各国政府がオープンデータに取り組む理由の一つとして、公共データを二次利用した新しいサービス市場を創造する目的がある。英国では、オープンデータを活用したビジネスを本格的に立ち上げるための行政機関Open Data Instituteを2011年11月に設立し、スタートアップ支援を開始している。また、日本では、経済産業省が「オープンデータ・ビジネス・コンペティション」⁽³⁾を2015年3月に主催するなど、ビジネス化に向けた活動や支援の機運が高まってきている。更に、東京メトロでは、電車の運行状況などを用いた「オープンデータ活用コンテスト」⁽⁴⁾を主催するなど、オープンデータ化の波は、民間企業にまで広がっている。

このような中、富士通研究所では、早くからオー

ペンデータの可能性や将来性に着目し「オープンデータの五つ星」と呼ばれるLinked Open Data (LOD)に関する技術を研究開発してきた。^{(5) - (10)}

本稿では、富士通研究所が中心となって公開・運営しているLOD検索サービス「LOD4ALL」や地域特性発見ツール「EvaCva（エヴァシーヴァ）」などを中心に、オープンデータのエコシステム創出に向けた富士通研究所の取組みを紹介する。

オープンデータの五つ星

オープンデータの二次利用を促進するには、ライセンス表示とデータ形式を整備する必要がある。ライセンス表示では、主として「クリエイティブ・コモンズ・ライセンス」⁽¹¹⁾が使用されることが多く、政府データカタログサイトでも採用されている。

一方、データ形式に関しては、Webの発明者であるTim B. Lee氏が「オープンデータの五つ星」⁽¹²⁾を提案している（表-1）。この格付けでは、コンピュータ処理しやすい形式かどうかで星の数が決められている。例えば「人が読む」ことを前提としたテキストなどの非構造化データ（PDFなど）は一つ星、表などの構造を持ったデータ（Excelなど）は二つ星、アプリケーションに依存しない形式（CSV、XMLなど）であれば三つ星のように定められている。

四つ星以上は、Web上でデータを公開するのに適した形式として、従来のWebページと同様に個々のデータがURL（正確にはURI）を持っており、RDF（Resource Description Framework）と呼ばれる形式で記述する。更に五つ星は、従来のWebページのように関連するデータにリンクを設定した「Linked Data」と呼ばれるものである。このLinked Data形式で公開されているデータの集合体がLODである。

表-1 オープンデータの五つ星

格付け		一つ星 ★	二つ星 ★★	三つ星 ★★★	四つ星 ★★★★	五つ星 ★★★★★
		非構造化データ 例：PDF	構造化データ 例：Excel	アプリケーション 非依存 例：CSV, XML	Web公開に適する 例：URI, RDF	関連データが リンク 例：Linked Data
データ数 2015年3月時点	日本：16 268	12 372 (76%)	3180 (19%)	713 (4%)	3 (0.02%)	0 (0%)
	英国：10 056	2711 (27%)	1820 (18%)	5127 (51%)	278 (3%)	120 (1%)

表-1は、2015年3月時点の英国および日本政府のデータカタログサイトで公開されているデータ数の比較についても示している。日本は英国と比べて公開されているデータの数が多いが、ほとんどが一つ星である。一方、英国は、多くのデータが三つ星で公開されていることが分かる。オープンデータ活動の進捗に伴い、より星の多い形式でデータ公開が進むと考えられる。

また、二つ星（Excel）や三つ星（CSV）のデータを四つ星（RDF）に変換するサービスも登場している。例えば、理化学研究所から一般社団法人化したリンクデータが運営するオープンデータ活用支援プラットフォーム「LinkData.org」⁽¹³⁾もそのサービスの一つであり、ExcelやCSVなどの表形式のデータをJSONやRDFに変換できる。このようなサービスと、富士通研究所が開発した「リンク自動付与技術」⁽⁹⁾を組み合わせることで、三つ星以下で公開されたデータを五つ星データに昇格させることができる。

LODのメリット：データの掛け合わせ

LODの最大の特徴は、関連データへのリンクを有していることであり、そのメリットとして「データの掛け合わせ」が容易になることが挙げられる。本章では、富士通研究所で試作した行政事業分析アプリケーションを用いて、そのメリットを紹介する。

内閣官房では、国の全ての事業を各府省自らが事後点検した「行政事業レビュー」⁽¹⁴⁾をExcel形式で公開している。この行政事業レビューのデータを利用すれば、どの企業がどの府省から多く受注

しているかを分析できる。

一方、LODの中にはDBpedia⁽¹⁵⁾と呼ばれるWikipediaを基に作成された多言語知識ベースがある。これは、Wikipediaの中のInfoboxと呼ばれるテーブル（通常ページ右上に表示される表形式のデータ）をLinked Data化し、公開しているものである。DBpediaでは、SPARQLと呼ばれる問合せ言語を用いて、ある企業の本社所在地や従業員数などを参照できる。技術的な解説は文献（7）を参照されたい。

ここで、例えば行政事業レビューに登場する「ある企業F社」に関し、F社が主要株主となっている企業をDBpediaに問い合わせることで、F社の関連企業の一覧を作成できる。この一覧を用いて、行政事業レビューを再集計すれば、F社単独の受注状況だけでなく、F社グループの受注状況も分析できるようになる。図-1はその具体例を示すものである。F社単独（左の円グラフ）では、S省やK省からの受注が多くM省からの受注は少ないが、F社グループ（右の円グラフ）ではM省からも多く受注していることが分かるようになる。

上記の例は、行政事業レビューとDBpediaに同時に表れる「企業」を切り口として、データを掛け合わせている。このように、データをLinked Data（五つ星）にすることで、Excelデータだけでは分からなかった事象も分析できるようになる。

今回は、LODの代表例としてDBpediaを用いた。State of the LOD Cloud⁽¹⁶⁾によれば、2014年8月時点で1000を超えるデータセットがLODとして公開されている。アイデア次第で、様々なデータの掛け合わせができるだろう。

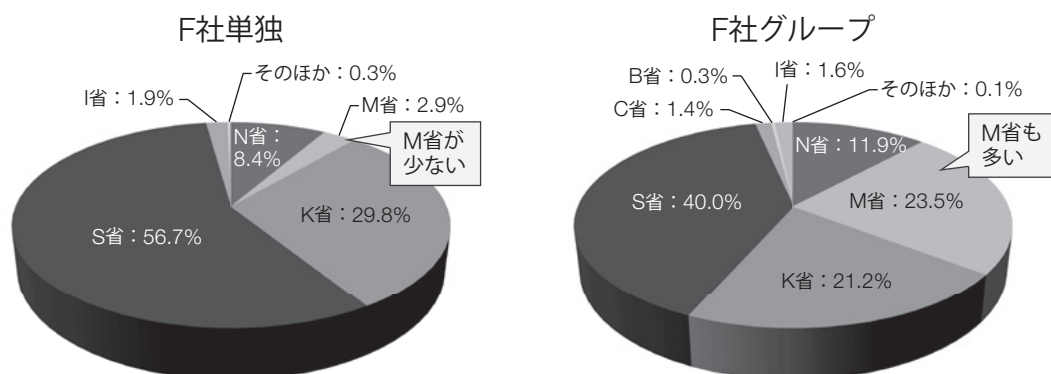


図-1 企業F社における府省別受注状況の分析

LOD検索サービス：LOD4ALL

富士通研究所では、LODの可能性にいち早く着目し、同分野において世界的な研究機関であるアイルランド国立大学ゴールウェイ校のINSIGHT Centre for Data Analyticsとの共同研究を2012年から開始した。この共同研究の一環として、Web上で公開されているLODを収集・格納し、複数のデータセットを一括して検索できるサービス「LOD4ALL」⁽¹⁷⁾を2014年1月より一般公開している。このサービスは、主に、データ利用者やアプリケーション開発者が、LODの中から利用したいデータセットを探すことを目的としたものである。前述のデータ変換・公開サービスであるLinkData.orgとも連携しており、データ作成者がデータを公開する場所としても利用できる。

また、2014年10月からはデータを取得するためのAPI (Application Programming Interface) を公開しており、本サービス上でアプリケーションを開発できる。図-2はAPIを利用して開発したアプリケーション例である。東急沿線情報サイト「と

くらく」が主催した「東急沿線データビジュアライゼーションコンテスト」⁽¹⁸⁾で公開されたデータを基に、総務省「政府統計の総合窓口」や国土交通省「国土数値情報」のデータを掛け合わせ、東急沿線の各駅を様々な指標で比較できる。

2015年3月時点で公開しているAPIは独自仕様を用いているが、2015年上期中にLODの標準APIであるSPARQLに対応する予定である。

地域特性発見ツール：EvaCva

本章では、LODを用いた本格的なアプリケーション事例として、地域の特性を見える化するWebツール「EvaCva」を紹介する。「EvaCva」は、公開されている統計データを活用して環境・経済・社会の観点から地域の特徴（強み・弱み）を明確化し、課題解決に向けた施策立案やビジネス展開を支援するツールである。特長を以下に示す。

- (1) 政府統計などの公開データから、地域特性を明らかにする上で必要と考えられる環境・経済・社会の47項目（表-2）を指標化し、偏差値評価でレーダーチャートや順位を表示する機能を設



図-2 LOD4ALLを利用して開発したアプリケーション例

表-2 EvaCvaにおける47指標の分類

ポリシー	評価要因	指標
経済	財政の健全性	・地方税歳入額 ・財政力指数 ・実質公債費比率
	産業の規模	・地域内総生産
	個人の所得	・課税対象所得 ・完全失業率
社会	社会インフラ	・病院数 ・公民館、図書館数 ・老人ホーム数 ・保育所数 ・下水道普及率 ・都市公園数
	社会サービス	・交通事故件数 ・刑法犯認知件数 ・火災件数 ・国民健康保険診療費 ・平均寿命 ・ごみのリサイクル比率 ・住居水準の充実度 ・一般行政職員数 ・教員当たり生徒数
	地域の活力	・人口自然増減率 ・人口社会増減率 ・空き家の数 ・住宅着工件数
環境	環境の資源	・林野面積比率 ・耕作放棄地比率 ・森林によるCO ₂ 吸収量
	環境の質	・河川水質 (pH, 有機物量, 浮遊物質, 溶存酸素量, 大腸菌群数) ・大気質 (オキシダント濃度, 浮遊粒子状物質濃度, CO濃度, SO ₂ 濃度)
	環境への負荷	・CO ₂ 排出量 (産業部門, 家庭部門, 民生業務部門, 運輸部門, 廃棄物分野) ・CO ₂ 総排出量 ・ごみ排出量 ・最終処分量 ・化学物質排出量によるヒト健康への影響度 ・化学物質排出量による生態系への影響度

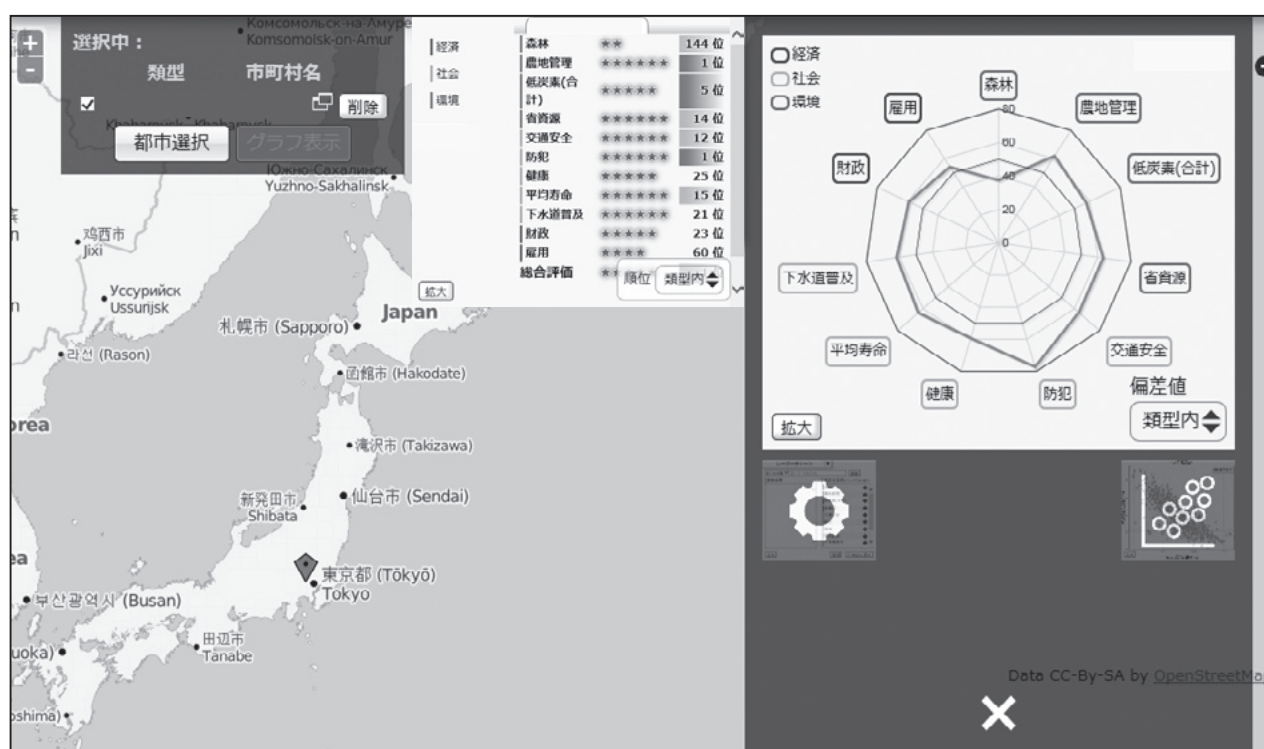


図-3 地域の分析事例

けている。

(2) 2014年4月1日時点での全国1742の市区町村を人口規模で10分類に類型化し、この分類に基づいて偏差値評価することで、類似自治体内での偏差値や順位を表示できる(図-3)。これにより、過疎化や少子高齢化の問題を抱えている小規模

な自治体と、人口集中の課題に直面している都市部の自治体を区別して比較できる。

(3) ユーザーが二つの指標(X軸・Y軸)を自由に選択してグラフ化できる。図-4は、X軸に15歳未満人口比率(%), Y軸に65歳以上人口比率(%)として少子高齢化の傾向を分析した事例である。

例えば、この図で示したA市の人口構成は、15歳未満の人口割合が小さく、65歳以上の人口割合が大きくなっており、少子高齢化の傾向が進んでいる状況が分かる。このように、2軸の相関グラフから傾向分析できるとともに、ユーザーが指定した自治体と類似した自治体の抽出も可能となる。

(4) LOD技術の適用し、インターネット上のオープンデータやユーザーが持っているデータを掛け合わせて多角的に分析することにより、価値ある情報の創出が可能となる。

主なユーザーの「EvaCva」の利用シーンを挙げてみる。

- ・自治体

地域活性化に向けた計画策定や施策立案，地域

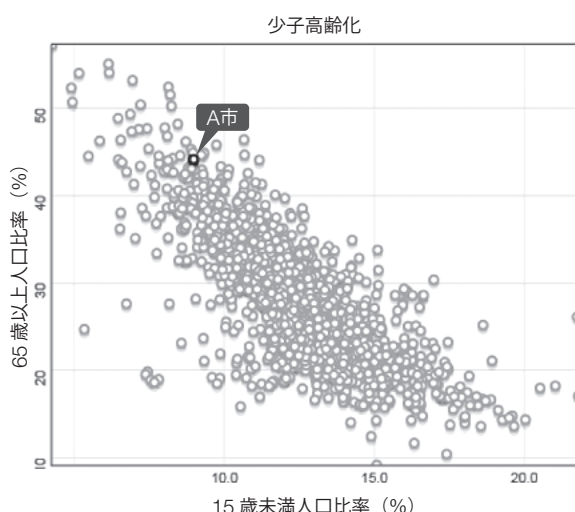


図-4 年少人口と老年人口の散布図

に適応した土地開発

- ・小売店や多店舗展開事業者

地域特性を見定めた出店計画や最適な商品展開

- ・地域住民や学術機関

地域活性化に向けた意見や提案，統計データに基づいた地域分析

「EvaCva」は、2015年12月までの予定で一般公開している⁽¹⁹⁾ 公開期間中に寄せられた要望を参考にして、今後も機能拡充を図っていく。

オープンデータのエコシステム

前章までに紹介したように、富士通研究所では研究成果を積極的に公開サービス化するとともに、外部団体のオープンデータ活動にも広く参画・協賛している。

一例として、国内最大級のオープンデータ・コンテストである「LODチャレンジ」⁽²⁰⁾ には、2011年の開始当初からスポンサーとして支援するとともに、2014年からはLODチャレンジに応募されたデータセットを格納した「LOD4ALL」を基盤提供している。LODチャレンジでは「データセット部門」「アプリケーション部門」などいくつかの部門があり、ある人が作成したデータセットを、別の人がアプリケーションで利用するといった分業が行われることもある。

また、LODには、データとアプリケーションの依存性を最小限にする効果がある。図-5に示すように、「EvaCva」のデータをLODとして公開することで、別の誰かが異なる派生アプリケーション①



図-5 オープンデータによるデータとアプリケーションの再利用

を作成できる（データの再利用）。更にデータに互換性があれば、派生アプリケーション①に別のデータを差し替えることで、更なる派生アプリケーション②を開発できる（アプリケーションの再利用）。実際、図-2で紹介した東急沿線駅の比較アプリケーションは、「LOD4ALL」のサンプルとして別途開発した自治体比較アプリケーション（EvaCvaデータの再利用）を再利用したものである。

このようなLODの効果は、多人数での分業を助け、より多くの派生物を産み出すことが可能である。富士通研究所の一連の活動は、こうした「オープンデータのエコシステム」の創出を目指したものである。

む す び

本稿では、オープンデータの五つ星と言われるLODに関し、LOD検索サービス「LOD4ALL」や地域特性発見ツール「EvaCva」などを中心に、富士通研究所の取り組みを紹介した。

GartnerのHype Cycle 2014⁽²¹⁾では、オープンデータは「Innovation Trigger」（黎明期）に位置づけられ、今後の利用拡大が期待されている。また、同レポートでは、企業がオープンデータを利用する際に、企業ごとに異なる競争優位性を確保する目的で、企業が所有するデータとの掛け合わせが必要と指摘されている。LODはまさしく「データの掛け合わせ」を容易にする技術であり、オープンデータだけでなく、企業情報システムへの適用も期待される。

参考文献

- (1) 内閣府：データカタログサイト。
<http://www.data.go.jp/>
- (2) Open Knowledge Foundation：Global Open Data Index。
<http://index.okfn.org/place/>
- (3) Idea.LinkData：オープンデータ・ビジネス・コンペティション。
<http://idea.linkdata.org/idea/idea1s800i>
- (4) インプレス：東京メトロ、オープンデータ活用コンテスト結果発表。
http://pc.watch.impress.co.jp/docs/news/20150221_689437.html
- (5) 井形伸之ほか：Linked Dataを用いた情報統合・活用技術。Fujitsu, Vol.64, No.5, p.464-470 (2013).
<http://img.jp.fujitsu.com/downloads/jp/jmag/vol64-5/paper01.pdf>
- (6) 西野文人：I-Discover—Linked Dataに基づく電子情報通信学会文献検索システム—。B-plus（電子情報通信学会通信ソサイエティマガジン）2013年夏号，電子情報通信学会，No.25, p.49-53, 2013.
- (7) 梅村 研ほか：テクノロジー最前線：活用広がるオープンデータ。日経SYSTEMS, No.259, p.56-61 (2014年11月号)。
- (8) 富士通研究所：オープンデータの活用革新！リンクが張られた公開データ（LOD：Linked Open Data）向け大規模データ格納・検索技術を開発。
<http://pr.fujitsu.com/jp/news/2013/04/3-1.html>
- (9) 富士通研究所：世界中で公開されているオープンデータへのリンクを自動的に付与する技術を開発。
<http://pr.fujitsu.com/jp/news/2014/01/16.html>
- (10) 富士通：Linked Open Data技術を適用した地域の特性を発見するツールを公開。
<http://pr.fujitsu.com/jp/news/2014/12/11.html>
- (11) クリエイティブ・コモンズ。
<http://creativecommons.jp/>
- (12) 5 Star Open Data。
<http://5stardata.info/>
- (13) LinkData.org：オープンデータ活用支援プラットフォーム。
<http://linkdata.org/>
- (14) 内閣官房：行政事業レビュー。
<http://www.cas.go.jp/jp/seisaku/gyoukaku/review.html>
- (15) DBpedia。
<http://dbpedia.org/>
- (16) Max Schmachtenberg et al.：State of the LOD Cloud 2014。
<http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>
- (17) LOD4ALL。
<http://lod4all.net/>
- (18) 東急沿線情報サイト とくらく：東急沿線データビジュアライゼーションコンテスト。
<http://www.tokyuensen.com/special/contest2014/>
- (19) オープンデータによる地域特性の発見EvaCva。

<http://evacva.net/>

(20) Linked Open Data Challenge Japan 2014.

<http://lod.sfc.keio.ac.jp/challenge2014/>

(21) Gartner : Hype Cycle for Business Intelligence and Analytics, 2014. p.27-28 (2014).

著者紹介



井形伸之 (いがた のぶゆき)

知識情報処理研究所
ナレッジシステムプロジェクト 所属
現在, Linked Dataを用いた情報統合・
活用基盤技術の研究に従事。



中澤克仁 (なかざわ かつひと)

R&D戦略本部
環境科学技術プロジェクト 所属
現在, 環境負荷評価技術および社会シ
ミュレーション技術の研究に従事。



桑 照宣 (くめ てるのぶ)

知識情報処理研究所
ナレッジシステムプロジェクト 所属
主に知識処理を中心とした研究を行っ
ており, 現在はLinked Data活用研究
に従事。



塩田哲義 (しおた てつよし)

R&D戦略本部
環境科学技術プロジェクト 所属
現在, オープンデータを用いた地域分
析技術の研究に従事。