

オープンデータ利活用に向けた 情報システム開発人材育成の動向

阿部 秀尚

要 旨

近年、OECD が発出した科学技術ポリシーに基づいて、欧米諸国のみならず、日本国内においても、公共の行政機関、団体などが保有する統計データのオープン化によるオープンデータ利活用の動きが広まってきている。また、CivicTech と呼ばれる情報関連技術を利用した各種の問題解決を草の根レベルで行う活動も、世界各地で広まってきている。このような背景のもと、オープンデータと組織内のデータとをうまく組み合わせ、企業活動に資する情報システム開発が求められている。本稿では、日本国内外の社会的・技術的背景、人材教育の動向について、解説する。

Abstract

In recent year, based on the OECD science and technical policy, The Principles and Guidelines for Access to Research Data from Public Funding, democratic countries including Japan began to utilize “open data” and open scheme innovation for activating their social and economic activities. Beside, CivicTech movements have spread to cities and towns in the countries all around the world for utilizing information technologies and the open data as non-commercial activity by the citizens in each region. With these backgrounds, many users in both of enterprises and governmental organizations latently desire to develop information systems, which utilize the open data by combining databases in each organization efficiently. In this article, the movements related to the open data innovation with information systems (IS) will be described from the following three viewpoints: social facts, technical trends, and educations for IS engineers.

1. はじめに

近年、OECD が発出した科学技術ポリシーに基づいて、欧米の民主主義国家を中心に、公共の行政機関、団体が保有するデータのオープン化が強力に推進されている。日本国内においても、2013 年 6 月 14 日付の閣議決定された「世界最先端 IT 国

家創造宣言」[1]に沿って、公共の行政機関、団体などが保有する統計データのオープン化が行われてきている。

しかし、公開されたオープンデータは、それだけでは意味を持たない数値や文字列の羅列であるため、適切な情報通信・情報処理技術と組み合わせることで、人々の役に立つ利活用が求められている。このようなオープンデータ利活用のため、各国政府はオープンなイノベーションによる情報技術産業の育成、学術団体を中心とした利活用の研究促進の動きが広がってきている。一方、CivicTech と呼ばれる情報関

Hidenao Abe

Faculty of Information and
Communications, Bunkyo University
文教大学 情報学部

[解説] 2016 年 9 月 11 日受付

© 情報システム学会

連技術を利用した各地の各種の問題解決を草の根レベルで行うための市民による活動も、世界各地の地方自治体レベルで広まってきている。オープンデータの増加と合わせ、CivicTech 活動によるオープンデータの利活用アイデアやシステムが増加しつつある。

以上のような背景のもと、情報システム開発を行う各企業に対しては、オープンデータを利用し、企業活動や自治体の活動に資する価値のある情報システム開発が求められようとしてきている。

本稿では、2章においてオープンデータの利活用が進む社会的状況について述べる。3章では、オープンデータを利活用するための技術について、主要な要素技術を開発する。さらに、4章では、各国の研究動向、人材教育の動向について、解説する。最後に、5章において、今後に向けた課題を提示し、本稿のまとめとする。

2. オープンデータを中心としたオープンイノベーションの社会的背景

オープンデータは、欧米の民主主義国家が推進してきた“開かれた政府”を背景に、各国の政府機関や地方自治体などの関連団体が法令や業務に基づいて収集したデータを公開しようとする活動が端緒となっている。初期からの取り組みとしては、英国における data.go.uk や米国における data.gov である。これらをポータルとして、様々な公開データに行き着くことができる。日本国内においては、data.go.jp 内で各府省の公開データを得ることができるが、都道府県や市町村の公開するオープン

データや総務省が公開する e-Stat に掲載された各種統計データへは、別々のサイトへアクセスする必要がある。

オープンデータは、組織内で発生するデータやインターネット上にあふれるデータ（総称して「ビッグデータ」などと呼ばれる）と混同されがちだが、適用されるライセンスや発生源などから区別できる。

図1は、Gurin が[2]内で定義した図の一部を公開した解説記事[3]で示された区分である。

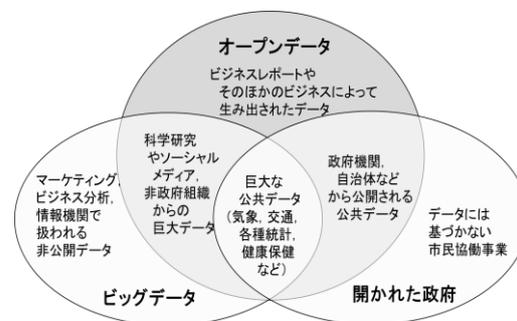


図1 オープンデータとビッグデータの関連と区分 ([3]より著者が翻訳)

オープンデータの発生主体は、政府の行政機関や地方自治体のみならず、民間企業や個人、市民団体が発生するデータである。これには、以下のような様々なオープンデータが含まれる。

- (ア) 科学研究，ソーシャルメディアからのデータ
- (イ) 巨大な公共データ (気象，各種統計調査，交通管制，健康保健調査などで発生)
- (ウ) 政府機関，自治体が生成する公共データ
- (エ) 民間企業が生成するビジネスレポートなどの公開データ

オープンデータは、機関や組織によって様々な形式で公開されるが、W3C ではその形式化度合を基準にガイドライン[4]を定め、統一した基準での公開や利用技術開発の促進を図っている。また、オープンデータの中には前述（ア）（イ）のデータのようにビッグデータとしての性質を持つものが存在する。このため、オープンデータを処理し、活用するシステムとして構築するための技術は、その多くがビッグデータを処理する要素技術と共通するところがある。

オープンデータは、原則としてクリエイティブ・コモンズの各ライセンス[5]に従って配布されることが多い。なかでも二次利用に制限を設けないCC-BYでの公開が推奨されている。これには、国民主権国家における政府機関や各自治体におけるデータが主権者たる国民のもの、という意識が強く働いていることが窺える。

オープンデータを利活用・普及するためのシステム開発は、Code for Yokohama（神奈川県横浜市）やCode for Sabae（福井県鯖江市）、Code for Kobe（兵庫県神戸市）といった”Code for”を冠し、地域での問題解決を一般市民の情報関連技術の技能を持ち寄って行うCivicTechと呼ばれる活動により、その多く行われている¹。最近では、Datathonと呼ぶ、オープンデータの活用アイデアを持ち寄り、簡単なシステムとして実現するためのイベントがCivicTech活動を中心に開催されている。また、アーバンデータチャレンジ[6]やLODチャレンジ[7]といった、産学連携や

学術機関を中心とした普及活動が日本国内で行われている。

一方、企業によるオープンデータを利用した商用の情報システム開発事例は、海外で多くみられ、企業内の（オープンではない）データとリンクしたシステムを開発し、顧客にサービスを提供することで価値を創造することで利益を得る仕組みづくりが求められている[8]。

3. オープンデータを活用するための基盤技術

オープンデータが想定するデータは、多くが表形式の統計データ（統計表）が念頭に置かれ、従来の情報システム上では関係データベース(RDB)に格納されている。ただし、オープンデータは表形式のデータだけではなく、個々の伝票を構造化したテキストデータ、さらにHTMLや通常のテキストデータから必要なデータ項目と値となるデータを取り出したものからも生成することができる。

データの利活用方法については、その多くが帳票入出力を念頭に置いた業務向け情報システムとは異なり、データに収録された「情報」を地図や図形を用いて視覚的に表現することが求められる。また、入力インタフェースについても、タブレット端末やスマートフォン（携帯型多機能端末）からテキストや画像の入力、タッチ操作によるポインティングによる入力方法が提供される。これは、利用者がこれまで情報システムを利用してこなかったような業務に従事するユーザ、および全く情報シス

を集約し、さらに別地域へ普及させるための活動を行っている。

¹ Code for Japan や Code for America など国家レベルでの団体は、各地域の活動

テム化が行われてこなかった活動でオープンデータを活用した情報システムを利用する市民のためである。

3.1 オープンデータの生成と公開技術

W3C が定めるガイドライン[4]のように、オープンデータは、機械（ソフトウェア・エージェント）可読の Web を築くための基盤となる「セマンティック Web」の考え方に強く関連する。そのため、CSV ファイルのデータ項目名や XML によるデータ記述のためのタグ語彙を統一する、あるいは同義であることが機械的な変換により判定可能である必要がある。また、値および文字列についても、単位系や文字コード体系を表記することができなければならない。

オープンデータは、タグ語彙と値の互換性表示が可能な RDF(Resource Definition Format)[8]によって公開されることが理想とされる。このようなオープンデータが 5 つ星とされる。RDF は XML を基盤として制定されたデータ規格であり、XML の名前空間を利用したタグ語彙の共通化、RDF によるスキーマ定義による値の書式・制限が明示できる、といった利点がある。タグ語彙の共通化には、Dublin Core[9]、共通語彙基盤[10]などを利用することで、複数のデータを結合可能にできる。また、RDF の値は文字列から成るリテラルと IRI(URI を拡張した識別子)によるリソースのどちらかで表されるため、複数のデータを結合することが可能となる。W3C の定義によると、RDF によって公開されたオープンデータを LOD (Linked Open Data)と呼ぶ。なお、RDF

は当初、XML を基本として規格が策定されたが、近年ではテキスト形式を拡張した Turtle 形式や JSON 形式に基づく JSON-LD[11]など軽量なデータ形式が用いられるようになってきている。

オープンデータの公開方法の類型を図 2 に示す。

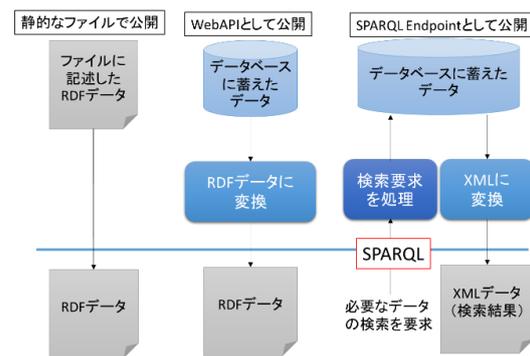


図 2 オープンデータの公開類型

データの公開は、RDF ファイルや CSV ファイルを直接 Web 上に公開する方式だけではなく、WebAPI としてデータ全体をデータベースから読みだして提供する方法、SPARQL による検索機能を付加した SPARQL Endpoint による公開方法がある[12]。基盤となる技術は HTTP を基本プロトコルとする Web であり、データベースからの公開には PHP, Java などによる Web プログラミング技術を要する。

3.2 オープンデータの利用技術

オープンデータを利用したシステムは、前述のように、従来の情報システムとは利用者層、利用する主機能が異なる情報システムに利用されることが多い。

以下に、LOD をはじめとするオープンデータを利用した情報システムを構築する際の基盤技術をおおまかに挙げる。

- A) オープンデータを情報として提示するための技術
- B) オープンデータから情報抽出技術
- C) オープンデータの結合・再構築技術

以上の基盤技術に加え、最も基盤的な技術はWeb(HTTP)を通じたデータの取得である。そのため、クライアント側ではHTTPクライアントとして機能する実装を行う。

オープンデータ活用のための情報提示技術

オープンデータを情報として視覚的にとらえるため、頻繁に用いられる機能が「地図上への情報表示」「グラフやチャート、表形式によるデータの表示」である。

地図上への情報表示では、Webを通して得たオープンデータに含まれる緯度・経度を用いて、Google Maps[13]やOpen Street Map[14]を利用した地図上にマーカーやボックスとしてオープンデータの内容が効果的に表示される。地図への表示は、地理情報システムとして体系的に論じられてきているため、これらの知見を基にオープンデータの活用を行うことにつなげることが可能である。

一方、グラフやチャートによる情報表示では、棒や円によるグラフ、折れ線チャートに加え、ヒートマップ、樹状図、鶏頭図など多様な情報視覚化手法を提供するD3.js[15]やGoogle Charts[16]などを用いて、効果的に情報が表示される。これらの情報視覚化は、地図による表示による地理情報システムと比較して、課題に合わせて適用されているのが現状である。このため、

体系的な情報視覚化論、および活用面におけるデータ中心ジャーナリズム論の確立が求められる。

さらに、情報の提示としては、視覚化以外にもマルチモーダルな方法によって、利用者にとって使いやすい形で情報を提示する必要がある。このため、自然言語処理による対話文生成や制御、ロボットによる提示など、これまでの情報システム開発に比べて多様な情報提示技術を組み合わせる必要がある[17]。

オープンデータからの情報抽出技術

位置情報を伴ったオープンデータでは、各インスタンスを地図上に配置・視覚化することで、利用者に気づきを与えることができる。しかし、表示粒度の変更に伴った表示の変更を行う場合を考えると、行政区域ごとの集計など基本的な統計処理が必要となってくる。さらに、ヒートマップによる確率密度の表示のような高度な表示を行うためには、確率密度推定による統計処理が必要である。

また、オープンデータの活用時に問題となるデータの連結においては、値の状態からRDFのプロパティ(述語)を推定するなど、従来はデータマイニングの分野で議論されてきた技術的な要素も必要となる。

このように、オープンデータから情報を抽出し、活用につなげるためには、統計学、機械学習による各種解析手法、および情報視覚化手法を目的に合わせて選定する技能が求められる。この技能は、データを中心に活用を行うデータサイエンティストとして重要なスキルセット[18]とも重なる部分が多い。

オープンデータの結合・再構築技術

オープンデータの活用において、LODとして公開されたオープンデータが持つ連結可能性が大きな役割を果たす。データの連結可能性は、従来のデータモデルによる分散データベース論において、透過性（分割透過性・分散透過性など）として議論されてきた。例えば、RDFをベースとしたLODでは、この分散データベースの透過性において、表形式データのカラム名にあたるプロパティ（述語、とも呼ぶ）に共通した語彙を与えることで、連結可能性を提供している。また、RDFによるスキーマ定義機能によって、値（リテラル）の記述形式や単位などが定義できることにより、複数のLODを連結し、結合することが可能となる。LODの連結を容易にするため、Dublin Coreをはじめ、共通語彙基盤のように各分野でのオブジェクトを記述するための共通化されたプロパティ語彙基盤の整備が進められている。

オープンデータを活用する情報システムは、利用者に情報を提示することが注目されるが、さらに重要な役割として、オープンデータを再生成することが期待される。特に、自治体や非営利目的でのオープンデータの利用では、情報の消費者によるデータの再生成が重要な価値の生成となる、と考えられる。そのため、オープンデータを提示されたインタフェースから、別のオープンデータの入力に円滑につながる入力インタフェースを用意することも、情報システム開発者には求められる[19]。

4. オープンデータを活用する情報

システム開発者育成の動向

オープンデータを利活用する情報システムの開発は、従来の情報システム開発と比較して、より利用者・利用場面に沿った利便性が重視される。さらに、高度な情報処理を利用者に意識させることなく、提示や入力受付を行う必要がある。このため、3章で示したような広範な基盤技術への理解が求められる。また、オープンデータの利活用にあたっては、オープンを前提とした高度な民主主義的な基盤思想があり、この考え方を理解し、商用利用を含む問題解決につなげていくことが重要である。

以上のような情報システム開発が可能な人材育成に向けて、日本国内では、公共に資する取り組みとして、セマンティックWebの普及を図る取り組みが学術研究分野で進められてきた[20]。現在、各大学における教育プログラムとしては、データベースの応用科目として、あるいはPBL型学習(Problem/Project Based Learning)の題材としてオープンデータを利活用するための情報システム開発が取り上げられ始めている[21]。産業界からも、オープンデータを利用した情報システム開発者育成への期待が見受けられるようになり[19]、オープンデータ利活用に向けた技術者向けの連続講座なども盛んに開催されている[22]。

一方、海外では社会的な問題解決をデータ解析技術によって行う「社会善」実現のためのデータサイエンス(Data Science for Social Good)を実現するためのシカゴ大学大学院での教育プログラム[23]をはじめ、いくつかの体系的な教育プログラムが開始されている。

さらに、学術的な国際会議の開催を契機として、CivicTech 活動と連携して、オープンデータの利活用に関する知見の共有とさらなる学術的な発展をめざす活動も試みられている[24].

オープンデータを利活用し、より良い社会を築く基盤として、オープンデータ自体の効率的な生成や利用しやすい情報提示システムの開発が求められる案件の増加に伴い、オープンデータを適切扱える情報システム開発者の需要は高まっていくものと考えられる。

5. おわりに

オープンデータを活用した情報システムの開発は、CivicTech の広まりとともに市民による自発的活動としては国内外問わず拡がりを見せてきているが、日本国内の企業による体系的な開発案件としては成り立ってきていない。オープンデータの利活用事例は、現在のところ、非営利による自発的なシステム開発が主流であるが、UNIX系オープンソース OS である Linux を利用した商業化サービスが可能であったように、決して商業化不可能ではない。

オープンデータを利活用するためには、データとして記録された事実に基づく冷静な判断による意思決定、社会により良い影響を自発的な行動で蓄積していく民主主義国家における市民としての態度の醸成、従来のコンピュータシステムを中心とした情報システム開発以外にも、より広範な社会的意義を含む情報システム開発が求められていくものと考えられる。

参考文献

- [1] 世界最先端 IT 国家創造宣言. (オンライン) (引用日: 2016 年 9 月 23 日.)
<http://www.itdashboard.go.jp/Achievement/index>.
- [2] J.Gurin. Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation. McGraw-Hill Education, 2014.
- [3] GurinJ. Big Data vs Open Data - Mapping It Out. (オンライン) (引用日: 2016 年 9 月 23 日.)
<http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/>.
- [4] W3C. Publishing Open Government Data. (オンライン) (引用日: 2016 年 9 月 23 日.)
<https://www.w3.org/TR/gov-data/>.
- [5] クリエイティブ・コモンズ・ライセンスとは. クリエイティブ・コモンズ. (オンライン) (引用日: 2016 年 9 月 23 日.)
<http://creativecommons.jp/licenses/>.
- [6] アーバンデータチャレンジ. (オンライン) (引用日: 2016 年 9 月 23 日.)
<http://urbandata-challenge.jp/>.
- [7] LOD チャレンジ. Wikipedia. (オンライン) (引用日: 2016 年 9 月 23 日.)
<https://ja.wikipedia.org/wiki/LOD%E3%83%81%E3%83%A3%E3%83%AC%E3%83%B3%E3%82%B8>.
- [8] W3C. RDF/XML Syntax Specification (Revised). (オンライン) (引用日: 2016 年 9 月 23 日.)
<https://www.w3.org/TR/REC-rdf>.

- syntax/.
- [9] DCMI Metadata Terms. (オンライン)
(引用日: 2016年9月23日.)
<http://dublincore.org/documents/dcmi-terms/>.
- [10] IPA 独立行政法人 情報処理推進機構: 共通語彙基盤. (オンライン) (引用日: 2016年9月23日.)
<https://www.ipa.go.jp/osc/kyoutsugou/ikiban>.
- [11] JSON for Linking Data. (オンライン) (引用日: 2016年9月23日.)
<http://json-ld.org/>.
- [12] 加藤文彦. オープンデータ時代の標準 Web API SPARQL. インプレス R&D, 2015.
- [13] Google Maps API. (オンライン) (引用日: 2016年9月23日.)
<https://developers.google.com/maps>.
- [14] Open Street Map Foundation (OSMF). Open Street Map. (オンライン) (引用日: 2016年9月23日.)
<https://www.openstreetmap.org/>.
- [15] Data Driven Documents. (オンライン) (引用日: 2016年9月29日.)
<https://d3js.org/>.
- [16] Google Charts. (オンライン) (引用日: 2016年9月29日.)
<https://developers.google.com/chart>.
- [17] 菅陽哉, 森雄一郎, 森田武史, 山口高平. PRINTEPS を利用した小学校社会科教育実践. 第 29 回人工知能学会全国大会, 1I4-3, 2015.
- [18] データサイエンティスト協会. データサイエンティスト スキルチェックリスト. (オンライン) (引用日: 2016年9月23日.)
<http://www.datascientist.or.jp/common/docs/skillcheck.pdf>.
- [19] 経済産業省. 平成 26 年度電子経済産業省構築事業 (オープンデータを活用したビジネス化の支援に関する調査研究) 調査報告書. (オンライン) (引用日: 2016年9月23日.)
http://www.meti.go.jp/meti_lib/report/2015fy/000482.pdf.
- [20] 森田武史, 山口高平. Linked Data を利用した情報統合. 人工知能学会誌, Vol. 27, No.2, pp.189-199. 2012, 第 27 巻.
- [21] PBL Summit 実行委員会. PBL Summit 2016. (オンライン) 2016年3月22日. (引用日: 2016年9月23日.) <http://pblsummit.jp/>.
- [22] リンクト・オープン・データ・イニシアティブ. LOD 連続講座. (オンライン) (引用日: 2016年9月23日.)
<http://linkedopendata.jp/?cat=17>.
- [23] The University of Chicago. Data Science for Social Good. (オンライン) (引用日: 2016年9月23日.)
<http://dssg.uchicago.edu/>.
- [24] ウェブを AI 化する?! セマンティック技術をテーマにした国際会議が秋に神戸で開催. WirelessWire News. (オンライン) 2016年3月16日. (引用日: 2016年9月23日.)
<https://wirelesswire.jp/2016/03/51232/>.