

# “『社会善』のためのデータサイエンス”をめぐる動向

阿部 秀尚†

Hidenao Abe†

あらまし 現在、欧米を中心に各国政府が進めている公共データの公開促進政策（オープンデータ化）と、これに呼応してより良い社会を情報通信技術の力によって実現しようとする研究活動、およびプログラミング技術を中心とした草の根運動が米国を中心に広がっている。本稿では、このような活動をデータという側面からみた“『社会善』のためのデータサイエンス”を中心に、その動向を概説する。

キーワード：社会善 (Social Good), データサイエンス, オープンデータ, ビッグデータ

## 1. はじめに

近年、欧米を中心としたオープンデータを用いた新たな価値創造が市民活動とともに広がりを見せてきている。オープンデータを用いた市民活動は、社会の問題解決に情報技術を利用し、“社会善”(Social Good)を実現しようとする民主的な取り組みと考えられる。このような社会の動向は、OECDの技術ポリシー<sup>1)</sup>の下、先進各国で公共機関の収集したデータを公開する活動と連動している。

一方、企業内においては、移動型携帯端末やIC内蔵型カードの普及により、ユーザの移動に関するデータが大量に蓄積されている。さらに、SNSの普及によって、多くのテキストデータが時々刻々と生成されている。このように、時間経過に従って高頻度に、絶え間なく蓄積可能なデータは“ビッグデータ”と呼ばれ、データの収集や分析に基づく企業活動の普及を大きなビジネス機会とみて、多くのIT関連企業が注目している。

本稿では、ビッグデータを利益追求型のビジネスとしていく方向性と区別し、より良い公共生活を得ようとする“社会善”を実現するためのオープンデータ活用の普及について概説する。

## 2. オープンデータを取り巻く現状

オープンデータは、欧米の民主主義国家における“開かれた政府”を推進してきた背景を基に、各国の政府機関や自治体が収集したデータを公開しようとする活動として公開が進められている。日本国内においては、ビッグデータ

ビジネスの普及と共に認知されるようになったため、その取り組みの多くが区分されず、用語の定義を十分せずに用いられている傾向にある。

用語の混同を避けるため、Gurinのブログ記事<sup>2)</sup>に基づき、図1のようにそれぞれの区分を定義する。

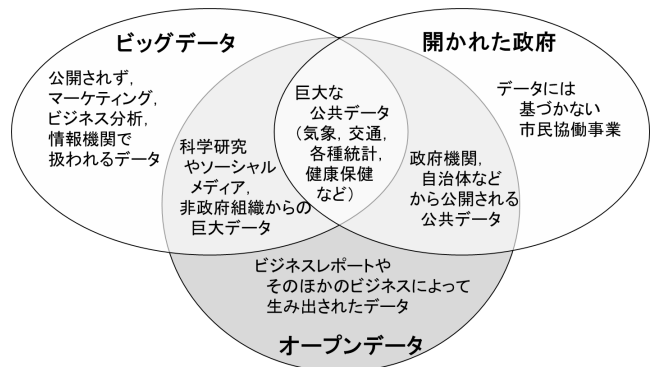


図1 オープンデータ、ビッグデータ、開かれた政府の関係とそれぞれの領域で発生するデータ。

オープンデータの発生主体は、政府機関や自治体のみならず、民間企業や個人が発生する公開されたデータも含まれ、以下のようなデータがオープンデータとして扱われる。

- (1) 科学調査、ソーシャルメディアからのデータ
- (2) 巨大な公共データ（気象、各種統計調査、交通管制、健康保健などで発生）
- (3) 政府機関、自治体が随時生成する公共データ
- (4) 民間企業などが随時生成する一般向けのデータ

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を定め<sup>3)</sup>、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上記1. 2. 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

2014年7月7日受付

〒253-8550 神奈川県茅ヶ崎市行谷1100

hidenao@shonan.bunkyo.ac.jp

† Graduate School of Information and Communications,

Bunkyo University

1100 Namegaya, Chigasaki, Kanagawa 253-8550, Japan

あり、このようなオープンデータの活用にはビッグデータの分析手法が必要とされる。

一方、ビッグデータという用語の浸透と共に、企業の間ではデータの活用を扱うデータサイエンティストと称する人材が求められている\*。データサイエンティストに求められるスキルセットは、現場の業務知識に加え統計学、機械学習、データ視覚化など多岐にわたるが、これらはオープンデータを扱う上でも重要なスキルセットとなっている。

### 3. 欧米での取り組み

「開かれた政府」推進の下、欧米ではこれまで多くの研究プロジェクトが活発に実施され、オープンデータをはじめとするオープンな基準<sup>4)</sup>での多くの活動が推奨されてきた。このような個人個人の善意に基づく自発的な活動の一つとして、地域社会の課題をオープンデータなどの情報関連技術を用いて解決しようとする Code for America\*\*をはじめとする Civic Tech 運動が各地に広まっている。

学術研究の分野においては、データマイニングと知識発見を中心に扱う ACM SIGKDD において、ここ数年、データ活用の公共的な役割が大きく取り上げられている。特に、2014年度の KDD2014(2014年8月24日~27日)\*\*\*では、会議の全体テーマを“Data Science for Social Good”としている。このほか、セマンティック Web に関する国際会議の ISWC をはじめ、データマイニングを中心に扱う国際会議である PKDD/ECML においても、オープンデータ活用に向けての国際会議、並びに国際ワークショップが開催されている。

人材育成においては、データサイエンスを扱う実践的な高度人材育成プログラムが各大学の大学院教育として実施され、多くの学生の注目を集めている。中でも、シカゴ大学の大学院課程における Data Science for Social Good プログラム<sup>5)</sup>は、データサイエンスによる“社会善”の実現を目指す人材育成を掲げる特徴的なプログラムである。

### 4. 日本国内での取り組み

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<sup>6)</sup>。しかしながら、データのオープン化に関する活動の多くはデータ連携の問題解決という学術研究としての側面が大きかった。

日本政府の取り組みは、OECD の技術ポリシーに従い、2013年6月14日付の閣議決定のもと、政府を挙げてオープンデータの利活用を促進する取り組みや基盤の整備を行っている<sup>7)</sup>。これまで、従来の統計データの電子化は総務省統計局、技術的基盤整備は経済産業省が中心に行ってきたが、内閣府を中心とした省庁横断的な体制が整いつつある。

一方、学術機関や民間の公共データ利活用については、リンクト・オープンデータ・イニシアティブ<sup>4)</sup>を中心に学術研究発表や市民向け講座の開催を通じて、その技術的基盤を広めつつある。また、Linked Open Data (LOD) チャレンジ<sup>5)</sup>が毎年開催され、年々多くの参加者が集まるイベントとなっている。

大学における実践的人材の育成については、「セマンティック Web」を表す科目において、オープンデータを扱う基礎となる RDF や SPARQL などの基本技術が扱われている。本学においても、3年次「プロジェクト演習」において、オープンデータを利活用しようとするプロジェクトが実施され、地方自治体のオープンデータ化を促進するプロジェクトは LOD チャレンジに参加するに至った<sup>6)</sup>。さらに、2015年度以降は、情報学部情報システム学科専門科目である「データベース応用」において、オープンデータを扱う実装を含めたシステム開発が行える実践的な人材育成が行われる。

### 5. おわりに

本稿では、オープンデータの利活用により、より良い社会の構築に繋げる“社会善”のためのデータサイエンスの取り組みについて、国内外の状況を踏まえて概説した。

今後は、オープンデータ化の技術的側面と社会的意義を理解する人材が多く輩出され、より多くの公共データの開示や連携基盤が整備されていくことが期待される。

### 【文 献】

- 1) OECD Principles and Guidelines for Access to Research Data from Public Funding: <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm> (Accessed 5th, July 2014).
- 2) Big Data vs Open Data? Mapping It Out: <http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/> (Accessed 5th, July 2014).
- 3) Publishing Open Government Data: <http://www.w3.org/TR/gov-data/> (Accessed 5th, July 2014).
- 4) The OpenStand Principle: <http://open-stand.org/principles/> (Accessed 6th, July 2014).
- 5) Data Science for Social Good: <http://dssg.io/> (Accessed 5th, July 2014).
- 6) 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).
- 7) IT DASHBOARD: <http://www.itdashboard.go.jp/> (Accessed 6th, July 2014).

あべ ひでなお  
**阿部 秀尚** 2004年9月静岡大学大学院理工学研究科博士後期課程修了。博士(工学)。2004年10月島根大学医学部医学科助教。2012年4月より文教大学情報学部講師。2013年4月より大学院情報学研究科情報学部専攻講師を兼任。専門は、データマイニングおよび知識工学。本情報学研究科では「情報システム特論」を担当。

\* <http://itpro.nikkeibp.co.jp/article/Interview/20130403/468356/>

\*\* <http://codeforamerica.org/>

\*\*\* <http://www.kdd.org/kdd2014/>

<sup>4)</sup> <http://linkedopendata.jp/>

<sup>5)</sup> <https://www.facebook.com/LOD.challenge.Japan>

<sup>6)</sup> [http://lod.sfc.keio.ac.jp/challenge2013/show\\_status.php?id=b007](http://lod.sfc.keio.ac.jp/challenge2013/show_status.php?id=b007),  
[http://lod.sfc.keio.ac.jp/challenge2013/show\\_status.php?id=b008](http://lod.sfc.keio.ac.jp/challenge2013/show_status.php?id=b008)