

## DBpedia Japaneseにおける利活用状況の把握と分析

## Use case study of DBpedia Japanese

加藤 文彦<sup>\*1</sup> 大向 一輝<sup>\*2</sup> 武田 英明<sup>\*2</sup>  
 Fumihiro Kato Ikki Ohmukai Hideaki Takeda

<sup>\*1</sup>情報・システム研究機構      <sup>\*2</sup>国立情報学研究所  
 Research Organization of Information and Systems      National Institute of Informatics

Opening up research datasets has been one of important topics in academic fields. One of common issues of this topic is that dataset providers would like to know the effect of opening and sharing their datasets. We tackle with the issue to evaluate the usage of our DBpedia Japanese which is one of the largest linked open datasets in Japanese. In the result, 78% of research related articles using DBpedia Japanese have been written by other researchers after publishing DBpedia Japanese in 2012. And 25 applications using DBpedia Japanese and 28 datasets linked to DBpedia Japanese have been created so far.

## 1. はじめに

オープンデータや Linked Data の発展により、行政をはじめとした様々な分野でデータの公開が進められている。学術においても、オープンサイエンスの一環として研究データの公開共有について議論されるようになっており、今後は研究者にとってもデータを公開することは重要な行為になるであろう。一方で、データ作成や管理にはコストがかかることもあり、データ提供者にとってデータを公開することのメリットは何か、データ提供者はどのように評価されるのかといったことが大きな課題となっている。

データ提供者としての評価の一つは実際にどのようにデータが利活用されているのかであるが、その実態を把握するのは困難である。研究分野においては、ScientificData の Data Descriptor<sup>\*1</sup> や Semantic Web Journal の Dataset Description paper<sup>\*2</sup> のように、データセット自体を業績として参照可能とする仕組みを導入しようとしており、それらによってデータ利活用の把握が部分的に可能になると考えられるが、まだこれからである。

筆者らは DBpedia の日本語版である DBpedia Japanese を 2012 年より公開している [Kato 13]。当時はまだ日本語で公開されている Linked Data が少なかったため、日本においても Linked Data の研究開発を促進するために公開を行った。主な目的は、DBpedia Japanese 自体を汎用的な Linked Data として様々な形で利活用してもらうことと、他の日本語のデータセットからとりあえずリンク可能な場所を提供するということがであった。

そこで本稿では、DBpedia Japanese の公開が具体的に研究コミュニティ及び一般にどのような影響を与えたのかを把握するために、アプリケーション・データセット・研究の 3 種類において、独自に利活用の現状を調査した。調査は人手で実施しているため、全ての事例を網羅しているわけではない。なお、本稿の調査結果をまとめたデータは、別途 [Kato 16] として GitHub に公開しているので、詳細はそちらも参照されたい。

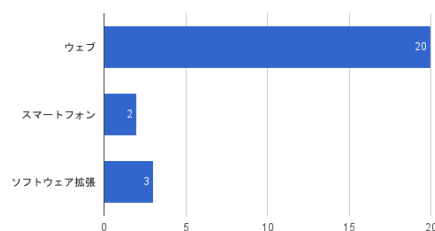


図 1: アプリケーション: プラットフォーム

## 2. アプリケーション

何らかの形で DBpedia Japanese を利用しているアプリケーションの総数は、執筆時点で 25 件であった (表 1)。アプリケーションは、分野を限定しない汎用的なツールやサービスと、特定分野における検索や可視化に大別できる。汎用が 11 件、特定分野利用が 14 件とほぼ半々である。特定分野の利用においては、スポーツや音楽、歴史、生物等、作者の目的に応じた様々な分野に適用されている。これは DBpedia Japanese 自体が汎用的なデータセットであることが活かされていると言える。

アプリケーションプラットフォームの殆どはウェブアプリとしてであり、スマートフォンのネイティブアプリは 2 件のみであった (図 1)。特徴的なのは、既存ソフトウェアの拡張プラグインとして実装しているのが 3 件あることである。それぞれの元となるソフトウェアは、GIS ソフトウェアである QGIS、ブラウザの Google Chrome、ブログシステムの WordPress とバラバラであるが、既存システムに取り込むための方法として他のシステムでも応用可能であり、注目すべきである。

利用方法については、殆どが DBpedia Japanese の SPARQL エンドポイントをウェブ API の一種として利用しているようである。一方で LOD4ALL のように、独自にデータセットのダンプをシステムに取り込んで利用している例もある。

連絡先: 加藤 文彦, 情報・システム研究機構, 〒101-8430 東京都千代田区一ツ橋 2-1-2 国立情報学研究所, fumi@nii.ac.jp

\*1 <http://www.nature.com/sdata/>

\*2 <http://www.semantic-web-journal.net/accepted-datasets>

アプリケーション名	対象	分野	アプリ実装	LODC 年度	LODC 部門	LODC 賞
QGIS プラグイン PediaLayer	汎用	地理	ソフトウェア拡張	2015	基盤技術	2015 基盤技術部門 優秀賞
なんでもリンク	汎用	ツール	ウェブ	2015	基盤技術	
ここベディア	汎用	地理	ウェブ	2015	アプリケーション	
隣の町から町へ列島プチ横断クイズ	特定	ゲーム	ウェブ	2015	アプリケーション	2015 アプリケーション部門 優秀賞
大河航役	特定	メディア	ウェブ	2015	アプリケーション	2015 アプリケーション部門 最優秀賞
放置自転車マップ	特定	生活	ウェブ	2015	アプリケーション	2015 データセット部門 最優秀賞
SakeLien: 日本酒とワインをつなぐ	特定	生活	スマートフォン	2015	アプリケーション	
WordPress Plugin	汎用	ツール	ソフトウェア拡張	2015	アプリケーション	
MusicSPARQL	特定	音楽	ウェブ	2014	アプリケーション	2014 学生奨励賞(大学部門)
ウィキ町史ビューアー	特定	歴史	ウェブ	2014	アプリケーション	2014 アイディア部門 最優秀賞
SPARQL Timeliner	汎用	可視化	ウェブ	2013	基盤技術	2013 基盤技術部門 優秀賞、公益財団法人横浜市芸術文化振興財団賞
LODLinker	汎用	検索	ウェブ	2013	基盤技術	
プロ野球因縁サーチ	特定	スポーツ	ウェブ	2013	アプリケーション	
日本の大学	特定	教育	ウェブ	2013	アプリケーション	
n 次の隣たり	汎用	検索	ウェブ	2013	アプリケーション	
DBpedia Japanese 簡単サーチ	汎用	検索	ウェブ	2013	アプリケーション	
疾患コンパス	特定	医療	ウェブ	2013	アプリケーション	2013 アプリケーション部門 優秀賞
ハマ経クロニクル	特定	メディア	ウェブ	2013	アプリケーション	
電子書籍リマインダー	特定	出版	ソフトウェア拡張	2012	アプリケーション	
みんなで創る音声アシスタント「花之声」	特定	生物	スマートフォン	2012	アプリケーション	2012 音声アシスタント賞
DashSearch LD	汎用	検索	ウェブ	2011	アプリケーション	2011 ビジュアライゼーション賞
Wikipedia Category Consistency Checker	汎用	検索	ウェブ			
LOD4ALL	汎用	検索	ウェブ			
バイオミメティクス・オントロジーによるキーワード探索システム	特定	生物	ウェブ			
東日本大震災アーカイブ Fukushima	特定	防災	ウェブ			

表 1: DBpedia Japanese を利用しているアプリケーション

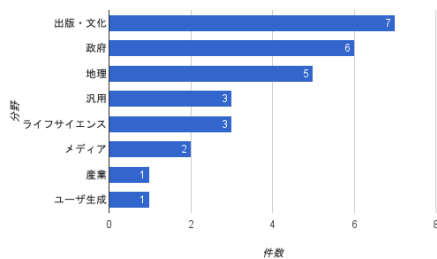


図 2: データセット: 分野

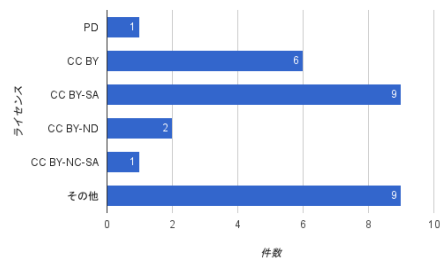


図 3: データセット: ライセンス

### 3. データセット

Linked Data の特徴として、データが互いにリンクすることによってデータセットそのものに価値がでるといえる点が挙げられる。そのため、DBpedia Japanese にリンクしているデータセットを評価することが DBpedia Japanese を評価する上で重要である。

本稿では、何らかの形で DBpedia Japanese の事物 IRI (例: <http://ja.dbpedia.org/resource/葛飾北斎>) にリンクしているデータセットのことを、DBpedia Japanese を利活用しているデータセットとする。つまり、DBpedia Japanese の HTML ページ (例: <http://ja.dbpedia.org/page/葛飾北斎>) にのみリンクしているようなデータセットは除外した。RDF のダンプファイルや SPARQL エンドポイントだけ公開しているデータセットのように、Linked Data の原則 [Tim 06] に必ずしも従っていないものも含んでいる。また、リンク数が少ないといった日本語 Linked Data Cloud の独自基準 [加藤 14] を満たしていないものも含んでいる。

DBpedia Japanese へリンクしているデータセットを調査したところ、総数は 28 件であった (表 2)。各データセットの分野については、データの中身に応じて日本語 Linked Data Cloud で用いている分野を筆者らが独自に判断して割り振っている。分野別で集計した結果が図 2 である。これから、出版・文化、政府、地理の順に多いことがわかる。

データセットのライセンスについて集計した結果が図 3 である。ライセンスが明示的に付与されているもので最多は CC

BY-SA となっている。また、いわゆるオープンライセンスとみなされているもの (Public Domain, CC BY, CC BY-SA) の採用数は 16 であり、全体の 57% であった。また、ライセンスを明示しているものの中では 84% がオープンライセンスであった。DBpedia Japanese にリンクするデータセットは比較的オープンライセンスを採用している傾向にある。

### 4. LOD チャレンジ

利活用事例を増加させる上で重要なのが、コンテスト等のイベント開催である。DBpedia Japanese については、LODAC プロジェクトが Linked Open Data チャレンジ Japan [乙守 12, 乙守 15] (以下、LOD チャレンジ) のデータ提供パートナーとして関わっている。2. 節と 3. 節で述べた各利活用事例の応募状況を調査すると、アプリケーションは 21 件 (アプリケーション総数の 84.0%)、データセットは 21 件 (データセット総数の 75.0%) が、LOD チャレンジの 1 部門あるいは複数の部門に応募をしていたことがわかった。アプリケーション部門とデータセット部門というように、同年度に複数の部門に応募している場合は、合わせて 1 件として数えている。

LOD チャレンジの応募年度毎に、アプリケーションとデータセットについて授賞した作品とそれ以外の作品で集計したのが図 4 と図 5 である。データセットについては、2011 年度と 2012 年度で 2 件の重複応募が存在する。そのため、図 5 ではそれぞれの年度で 1 を足しており、計 23 件となっている。また、複数部門に応募して、別部門で授賞したものについても授

データセット名	分類	ライセンス	LODC 年度	LODC 部門	LODC 賞
放置自転車 LOD	ユーザ生成	CC BY	2015	データセット	2015 データセット部門最優秀賞
大河配役 LOD	メディア	CC BY-SA	2015	データセット	2015 アプリケーション部門 最優秀賞
日本酒&酒蔵のデータ	地理	CC BY	2015	データセット	
RouteFinder 設置位置情報	地理	CC BY	2015	データセット	
NISTEP 大学・公的機関名辞書 LOD	政府	CC BY	2015	データセット	
GeoNames.jp	地理	CC BY-SA	2014	基盤技術	2014 基盤技術部門 最優秀賞
When.exe	汎用	CC BY-ND	2014	データセット	2014 データセット部門 優秀賞
Open Park Yokohama	政府	CC BY-SA	2014	データセット	
動物園 LOD	出版・文化	CC BY-SA	2014	データセット	
GeoLOD	地理	-	2013	基盤技術	2013 基盤技術部門 最優秀賞
東日本大震災アーカイブ Fukushima	政府	-	2013	データセット	
議員 LOD	政府	CC BY-SA	2013	データセット	
NHK 映像マップみちしる LOD	メディア	-	2013	データセット	
ねじ LOD	産業	CC BY-SA	2013	データセット	2013 データセット部門 最優秀賞
京都国際マンガミュージアム書誌情報 LOD	出版・文化	CC BY-NC-SA	2012	データセット	
Open DATA METI LOD	政府	-	2012	データセット	2012 オープンデータ推進賞
青空文庫 Linked Open Data	出版・文化	CC BY-SA	2012	データセット	2012 学生奨励賞
日本語 Wikipedia オントロジー	汎用	CC BY-SA	2011,2012	データセット	2012 データセット部門 優秀賞
LSJ: Location Site of Japanimation	地理	CC BY	2011,2012	データセット	2011 データセット部門 優秀賞, 2012 goo 賞
LODAC Species	ライフサイエンス	一部 CC BY-SA	2011	データセット	
ヨコハマ・アート・LOD	出版・文化	CC BY	2011	データセット	2011 オープン・ガバメント賞
LODAC Museum	出版・文化	-	-	-	-
ライフサイエンス辞書	ライフサイエンス	CC BY-ND	-	-	-
WordNet-ja	汎用	-	-	-	-
環境リポジトリプロトタイプシステム	ライフサイエンス	-	-	-	-
Evacva	政府	-	-	-	-
図書館及び関連組織のための国際標準識別子 (ISIL)	出版・文化	PD	-	-	-

表 2: DBpedia Japanese へリンクしているデータセット

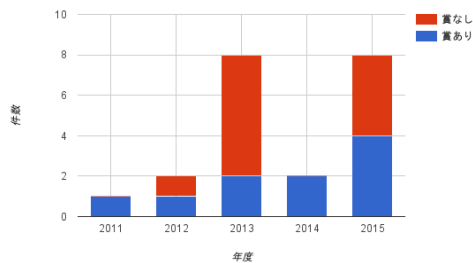


図 4: LOD チャレンジ: アプリケーション

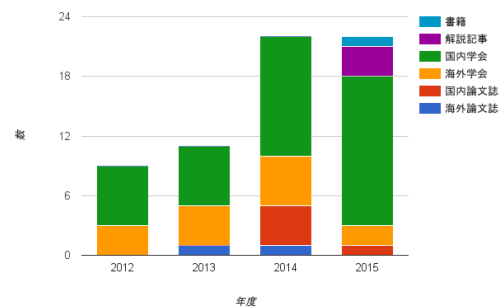


図 6: 研究活動での利活用

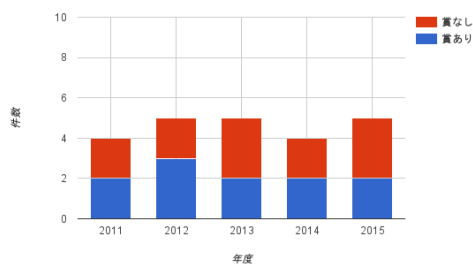


図 5: LOD チャレンジ: データセット

賞ありに含めている。

執筆時点で DBpedia Japanese を利活用しているものを挙げているため、LOD チャレンジの応募時点では DBpedia Japanese を用いていないものも含んでいる。DBpedia Japanese の公開が 2012 年であるのに 2011 年度の応募作品があるのは、DBpedia Japanese の公開後に対応したものと考えられる。

これらの図によると、アプリケーションとデータセット共に毎年数件授賞をしていることがわかる。また、授賞の中身を見ると、各部門の最優秀賞や優秀賞を授賞していることが多く、審査員特別賞等は少数である。アプリケーションは Linked

Data のアプリケーションとして評価されており、データセットについても DBpedia Japanese にリンクしているということで Linked Data としてもよく出来ていると評価されていると考えられる。

## 5. 研究活動における利活用

本節では、DBpedia Japanese を利活用している論文、研究発表、書籍等を独自に収集した結果について述べる。本節における利活用とは、DBpedia Japanese の分析、DBpedia Japanese を用いたアプリケーションの作成、データセットからの DBpedia Japanese へのリンクの作成等を指し、関連研究等で言及しているだけのものは含まない。また、原稿中には DBpedia と記述しているが、実データやアプリケーションでは DBpedia Japanese を用いている場合は、利活用例として含んでいる。

研究活動における利活用を集計した結果が図 6 である。2012 年からの総数は 64 件であり、そのうち LODAC プロジェクトの関係者を含まない件数は 50 件 (78.1%) であった。また、年度を追う毎に、学会でも論文誌でも数が増加傾向にある。これらから、データを広く公開共有したことで、研究分野の発展に一定の寄与をしたといえる。

---

## 6. おわりに

本稿では、LOD データセットの公開共有が研究活動やその他の利活用においてどのような影響を与えるのか知るための一例として、著者らが公開している DBpedia Japanese の実態調査を報告した。日本語のデータセットにも関わらず、公開から3年半でアプリケーション・データセット・研究関連で数多くの利活用実績があった。研究関連では8割弱が第三者による利活用であった。そのため、データセットを広く公開共有することは研究コミュニティを推進していくために一定の効果があると考えられる。

## 謝辞

第36回セマンティックウェブとオントロジー研究会参加者の方々には、DBpedia 及び DBpedia Japanese 利活用例の頭出しにご協力頂き、心より感謝申し上げます。

## 参考文献

- [Tim 06] Tim, B: Linked Data, <https://www.w3.org/DesignIssues/LinkedData.html> (2006)
- [Schmachtenberg 14] Schmachtenberg, M, Bizer, C, Jentzsch, A, Cyganiak, R: Linking Open Data cloud digram 2014, <http://lod-cloud.net> (2014)
- [Kato 13] Kato, F, Takeda, H, Koide, S, Ohmukai, I: Building DBpedia Japanese and Linked Data Cloud in Japanese, 2013 Linked Data in Practice Workshop (2013)
- [加藤 14] 加藤 文彦, 武田 英明, 小出 誠二, 大向 一輝: 日本語 Linked Data Cloud の現状, 第28回人工知能学会全国大会 (2014)
- [Kato 16] Fumihiko Kato: dbpedia-japanese-usecases:1.0.1, Zenodo. doi:10.5281/zenodo.48479 (2016)
- [乙守 12] 乙守 信行, 長野 伸一, 佐藤 宏之, 萩野 達也: Linked Open Data チャレンジ 2011 を振り返って, 人工知能学会誌, Vol.27, No.5, pp.518-5266 (2012)
- [乙守 15] 乙守 信行, 中辻 真, 萩野 達也: オープンデータの普及促進を加速させるコンテストの開催 - LOD チャレンジ Japan の取り組み -, 人工知能, Vol.30, No.5, pp.598-604 (2015)