

特集 「Linked Data とセマンティック技術」

Linked Data と Web アーキテクチャ

— DBpedia Japanese を例に —

Linked Data and Web Architecture — In the Case of DBpedia Japanese —

加藤 文彦
Fumihito Kato

情報・システム研究機構
Research Organization of Information and Systems.
fumi@nii.ac.jp, <http://www.rois.ac.jp>

Keywords: linked data, web architecture, DBpedia, RDF, SPARQL.

1. はじめに

Web の開発者であるティム・バーナーズ＝リー氏が 2006 年に Linked Data [Berners-Lee 06] という文書をまとめて以来、世界中で Linked Data に則った形でデータ公開を行う Web サイトが立ち上がってきている。それぞれのデータが互いにリンクすることによって、“データの Web” とでもいべきものができつつある。W3C の Semantic Web Education and Outreach (SWEO) Interest Group^{*1} が始めた Linking Open Data Community Project^{*2} では、Linked Data として公開されているデータセットが何らかのリンクでつながっているという関係を概観できる The Linking Open Data cloud diagram (以下、LOD クラウド)^{*3} を 2007 年より継続的に作成・公開している。

図 1 は 2007 年 5 月 1 日に公開された、LOD クラウド

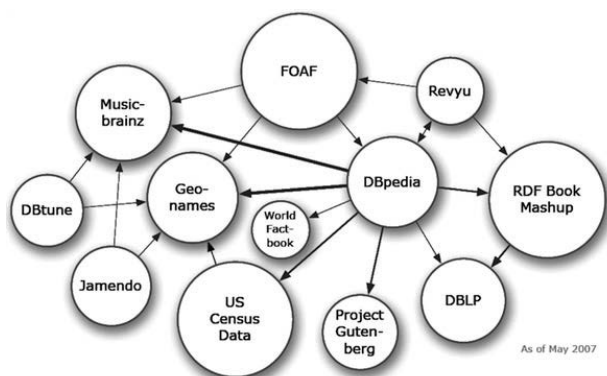


図 1 LOD クラウド 2007 年 5 月 1 日版

*1 SWEO 自体は 2008 年に終了しているが、コミュニティはメーリングリスト public-lod@w3.org として存続している。

*2 <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

*3 <http://lod-cloud.net>

ドの最初の版である。この図において、すでにリンクの中心の一つとなっているのが DBpedia [Lehmann 15] である。最新版の 2014 年 8 月 30 日版 (図 2) においては、DBpedia は名実ともに LOD クラウドの中心となっており、Linked Data の代表的なデータセットであるといえるだろう。

DBpedia は Wikipedia から構造化された多言語の知識を抽出することで、Linked Data として Web 上に再公開するコミュニティプロジェクトである。もともと Wikipedia という巨大な百科事典であるので、DBpedia は他のどんな分野のデータセットからでもリンクしやすい、汎用的なデータセットとなっている。多くのデータセットがとりあえず DBpedia のような汎用データセットにリンクすることで、それらをハブに他のデータセットにつながる機会が増えるというネットワーク効果を得られるようになってきている。それが LOD クラウドの中心に DBpedia がある理由である。

Wikipedia のコンテンツのライセンスはクリエイティブ・コモンズ表示-継承 3.0 非移植ライセンス (CC BY-SA) なので、DBpedia もそのライセンスを継承して、生成したデータを CC BY-SA のもとで公開している。RDF のダンプファイルとしては、執筆時点で 125 言語の Wikipedia を対象としたデータセットを DBpedia の Web サイト^{*4} で公開している。さらに DBpedia の Web サイトでは、主に英語版 Wikipedia を中心としたデータについて、Linked Data としてのデータ提供をしている。また、SPARQL による問合せを受け付けるための公開 SPARQL エンドポイントも提供している。

他の一部の言語については、有志の組織が各国語版の DBpedia サイトを公開しており、英語版と同様に Linked Data や SPARQL エンドポイントによるデータ提供をしている。執筆時点では 18 言語の DBpedia サイ

*4 <http://dbpedia.org>

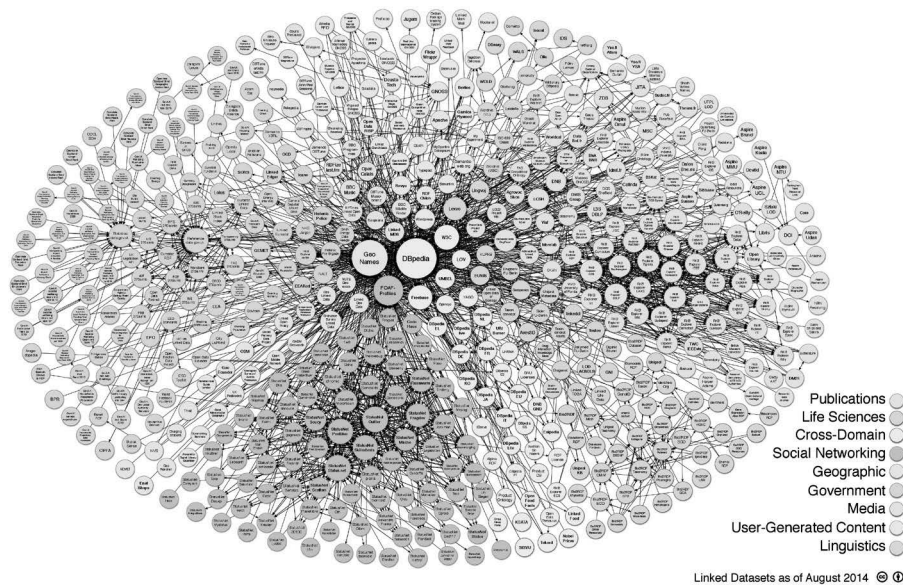


図2 Linking Open Data cloud diagram 2014 [Schmachtenberg 14a]

トが公開されており、著者らが2012年より運営しているDBpedia Japanese [Kato 13]^{*5}はその日本語版である。他に公開されている言語についてはInternational DBpedia chapters^{*6}を参照してほしい。

ここ数年で日本語においても、DBpedia Japaneseを中心としてRDFデータおよびLinked Dataの公開が徐々に広まってきている。しかし一方で、まだLODクラウドに入る基準(詳細は4.1節を参照)を満たしているものは少ない[加藤 14]。その理由の一つとしては、Linked DataやLODクラウド基準についての認知不足が考えられる。

そこで本稿では、DBpedia Japaneseの例を通して、Linked Dataとは何かとその構成要素について解説する。さらにLODクラウド基準についても解説することで、より多くの日本語のデータがLODクラウドに載ることができるようになることを期待したい。

2. Webアーキテクチャ

Linked DataはWeb技術でデータを公開・共有するためのベストプラクティスである。Linked Dataの特徴は、データへのアクセスに従来のWebのアーキテクチャを適用していることと、公開するデータのモデルとして統一のグラフデータモデルであるRDFを用いていることである。Linked Dataは、その4原則[Berners-Lee 06]として以下のように参照される。なお注意として、Linked Dataに関連する文書では歴史的経緯によりURI (Uniform Resource Identifier) [Berners-Lee 05]とその国際化版であるIRI (Internationalized Resource

Identifier) [Dürst 05]が混在して使われているが、本稿では参考文献のタイトルとして使われている箇所以外は、最新のRDF 1.1 [Cyganiak 14]やSPARQL 1.1 [Harris 13]に合わせてIRIで統一して表記しなおすこととする。

- (1) あらゆる事物 (thing) にIRIを付与すること
- (2) 誰でも事物の内容が確認できるように、HTTP IRIを用いること
- (3) IRIを参照したときは、標準の技術 (RDFやSPARQLなど) を使用して情報を利用できること
- (4) 情報には、より多くの事物を発見できるように、他の情報へのIRIリンクを含めること

この原則の背景には、より一般的なWebアーキテクチャ [Jacobs 04]がある。Linked Dataは、RDFデータを流通させる方法として従来のWebアーキテクチャを適用していると言い換えることができる。そのため、基本的なWebの仕組みをまず知る必要がある。

Webアーキテクチャにおいて、あらゆるものはリソース (resource) として抽象化されている。例えば、DBpediaのトップページや、日本語Wikipediaの夏というページといったWebページはすべてリソースである。また、おのおののページから読み込まれている画像などもリソースである。

そして、これらのリソースを識別して一意にアクセスするために、IRIが使われている。DBpediaのトップページは「<http://dbpedia.org>」というIRIで指し示されており、日本語Wikipediaの夏というページのIRIは「<https://ja.wikipedia.org/wiki/夏>」である。IRIがリソースを識別していること、そしてHTTP IRI (<http://> および <https://> で始まるIRI) を使用していることで、Webブラウザやクローラといったユーザーエージェントが、各リソースへHTTP経由で一意にア

*5 <http://ja.dbpedia.org>

*6 <http://dbpedia.org/about/language-chapters>

アクセスすることができる。

ある IRI で指し示されるリソースにアクセスすると、サーバから HTML 文書や画像ファイルなどが返ってくる。このプロセスのことを、参照解決 (dereference) という。また、リソースに対して、HTML 文書や画像などの実際にユーザエージェントとサーバの間で運ばれるオクテット列のことを表現 (representation) と呼ぶ。そのようなプロトコルのメッセージとして直接やり取りできる表現があるリソースのことを、情報リソース (information resource) とも呼んでいる。

多くの場合、情報リソースと表現は 1 対 1 の関係であるが、一つの情報リソースに対して複数の表現があってもよい。例えば東京の天気予報というリソースについて、日本語の HTML 文書と英語の HTML 文書が用意されているかもしれないし、印刷用の PDF ファイルも用意されているかもしれない。また、天気予報の晴れ画像について、PNG 画像と SVG 画像のように異なるファイル形式が用意されているかもしれない。ユーザエージェントは HTTP の内容折衝 (content negotiation) という機能で、日本語環境ならば日本語の HTML 文書を優先して取得するといった処理を常に行っている*7。

そして、取得した HTML 文書表現の中にリンクで IRI が指定されていることにより、ユーザはそれを辿って他のリソースにアクセスすることができる。Web では誰もが自由に関連するリソースへのリンクを記述して公開することができる。リンクが次々につくられていくことによって、Web はグローバルな情報空間として発展してきている。

まとめると、Web ページのような Web 上にある情報リソースの識別子として HTTP IRI があり、HTTP IRI を参照解決することで情報リソースの表現として HTML 文書などを取得できる。取得した HTML 文書には他の情報リソースへのリンクがあり、それを辿っていくことができる。これが基本的な Web の仕組みである。

3. Linked Data の原則

2 章で述べた Web アーキテクチャと Linked Data の原則を見比べると、実は内容がほとんど変わらないことに注目したい。しかし決定的に異なるのは、Linked Data の原則では事物を IRI で識別するということである。ここでいう事物とは、例えば葛飾北斎という人間や、東京駅という建物のように、具体的な実世界のオブジェクトから、トピックや知識体系のような抽象的な概念まで含んだものである。これらは情報リソースのように、直接メッセージとして運べる表現をもつわけではない。そのため、事物のことを“非情報リソース”として

明示的に情報リソースと区別する。本章では、DBpedia Japanese を例に、Linked Data の原則について詳細に紹介する。

3.1 原則 1: IRI による識別

DBpedia では、Wikipedia のエントリーが対象としているものを事物として扱う。DBpedia Japanese における“葛飾北斎”という人自身の IRI は「<http://ja.dbpedia.org/resource/葛飾北斎>」となっている。「<http://ja.dbpedia.org/resource/葛飾北斎>」に Web ブラウザでアクセスしてみると、「<http://ja.dbpedia.org/page/葛飾北斎>」という IRI にリダイレクトされて、図 3 のような画面が表示されるはずである。

Property	Value
dbpedia:owl:abstract	葛飾北斎 (かつしか ほうさい、葛飾北斎、宝暦10年9月23日 (1760年10月31日) ? - 嘉永2年4月18日 (1849年5月10日)) とは、江戸時代後期の浮世絵師。化政文化を代表する一人。代表作に『富嶽三十六景』や『北斎漫画』があり、世界的にも著名な画家である。
dbpedia:owl:birthDate	1760-10-31 (year+date)
dbpedia:owl:deathDate	1849-05-10 (year+date)
dbpedia:owl:deathPlace	dbpedia:江戸
dbpedia:owl:field	dbpedia:浮世絵

図 3 葛飾北斎 -DBpedia Japanese

ここで最も重要なのは、実世界の人である“葛飾北斎”を表す IRI (非情報リソース) と、“葛飾北斎”という人について書かれている HTML 表現の IRI (情報リソース) がそれぞれ明示的に区別されているということである。“葛飾北斎”という人自身に IRI があることが、Linked Data の原則の一つ目にある“あるゆる事物に IRI を付与”ということの意味である。

3.2 原則 2: HTTP IRI の利用

非情報リソースと情報リソースをともに IRI で識別することが重要だとして、次に、非情報リソースのための IRI の設計と、それを関連する情報リソースにどう結びつけるのかという問題がある。IRI はあらゆるものを世界中で一意に識別できるようにするために、さまざまな形式を受け付けるように設計されている。例えば、非情報リソースの IRI の例として、ISBN の URN がある。“Linked Data:Web をグローバルなデータ空間にする仕組み”という本そのものを指す IRI として、「<urn:isbn:978-4-7649-0427-9>」のように表すことが可能である [Hakala 01]。

しかし、URN は基本的に Web とは切り離された空間で定義されており、Web 上の関連する情報リソースに

*7 実際の表現選択は、内容折衝によらずにクッキーなどで実現している場合もよくある。

表1 DBpedia Japanese の IRI 設計

種別	情報・非情報	IRI パス部分	IRI 例
事物	非情報リソース	/resource/[エントリ名]	http://ja.dbpedia.org/resource/ 葛飾北斎
HTML 表現	情報リソース	/page/[エントリ名]	http://ja.dbpedia.org/page/ 葛飾北斎
Turtle 表現	情報リソース	/data/[エントリ名].ttl	http://ja.dbpedia.org/data/ 葛飾北斎 .ttl
RDF/XML 表現	情報リソース	/data/[エントリ名].rdf	http://ja.dbpedia.org/data/ 葛飾北斎 .rdf

結び付けて参照解決可能にするのが困難という問題がある。Web アーキテクチャに合わせて参照解決できるようにすることを考えるのであれば、非情報リソースにも HTTP IRI を用いるのが現実的な選択となる。これが、Linked Data の原則の二つ目にあたる。

3.3 原則3: IRI の参照解決

非情報リソースと情報リソースをともに HTTP IRI で表現することによって、非情報リソースの IRI を参照解決するときには何らかの情報リソースを HTTP で返すという手段を取ることができる。上の葛飾北斎の例では情報リソースとして HTML 文書を返していたが、ここで HTML 文書 (だけ) ではなく、RDF データを返せるようにしようというのが Linked Data の原則の三つ目である。非情報リソースと情報リソースについて、意味の整合性を保ちつつ実装するには、さまざまな方法が考えられるが、Cool URIs for Semantic Web [Sauermann 08] に、典型的な解決方法がまとめられている。DBpedia Japanese では、紹介されている解決方法のうち、「異なる文書への 303 IRI 転送」という手法を用いている。303 See Other [Fielding 99] は HTTP における返答のステータスコードの一つであり、Web サーバから代替のリダイレクト先を通知するときに使われる。303 が使われる理由は [httpRange-14](http://www.w3.org/2001/tag/group/track/issues/14)*8 および [HttpRedirections-57](http://www.w3.org/2001/tag/group/track/issues/57)*9 といった Web アーキテクチャの根幹問題として長年議論されており、将来別の方法がでてくることも十分にあり得る。

ある事物の IRI にアクセスされたとき、Web サーバが HTTP リクエストの Accept ヘッダに応じて適切な表現の IRI に 303 転送することで、事物と表現の IRI の分離をしつつ、リソースを Web 上で関連付けることができる。表1は DBpedia Japanese における IRI 設計の抜粋と葛飾北斎の例である。表現例として HTML, Turtle, RDF/XML を載せているが、他にも RDF/JSON など、多種の表現に対応している。IRI パス部分の [エントリ名] には、IRI 例のように Wikipedia のエントリ名がそのまま適用される。

また、葛飾北斎にアクセスしたときの、葛飾北斎に関する HTML 表現や Turtle 表現への転送を図4に示す。303 転送では、エージェントの Accept

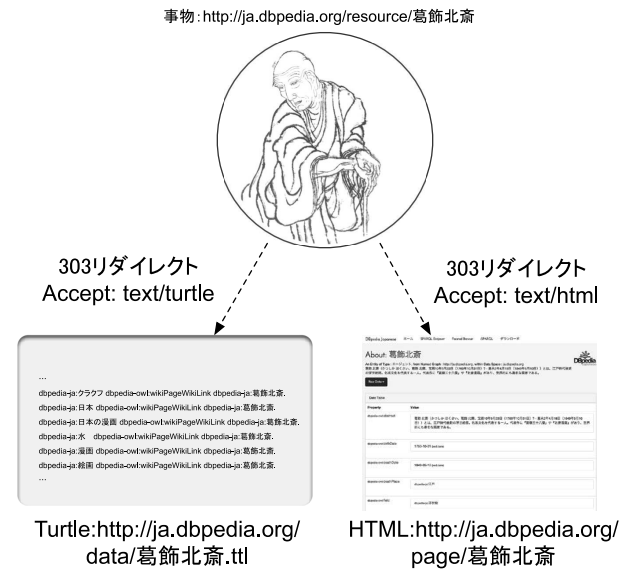


図4 DBpedia Japanese の 303 転送

ヘッダに合わせて、適切な表現が取得可能な IRI に転送する。DBpedia Japanese では RDF ストアとして Virtuoso のオープンソース版*10を採用しており、303 転送も DBpedia と同様に Virtuoso の拡張として実装している*11。

実装としては、Virtuoso 上で Accept ヘッダに対して IRI パス部分の書換えをするという方法をとっている。「[http://ja.dbpedia.org/resource/ 葛飾北斎](http://ja.dbpedia.org/resource/葛飾北斎)」に Web ブラウザのように Accept ヘッダで text/html を優先しているユーザエージェントがアクセスすると、303 で「[http://ja.dbpedia.org/page/ 葛飾北斎](http://ja.dbpedia.org/page/葛飾北斎)」に転送される。この HTML 表現は SPARQL の DESCRIBE の結果を HTML で表示している。葛飾北斎であれば、DESCRIBE <[http://ja.dbpedia.org/resource/ 葛飾北斎](http://ja.dbpedia.org/resource/葛飾北斎)>」である。

一方で、Linked Data のクローラのように Accept ヘッダで text/turtle を優先するユーザエージェントが「[http://ja.dbpedia.org/resource/ 葛飾北斎](http://ja.dbpedia.org/resource/葛飾北斎)」にアクセスすると、Turtle 表現の「[http://ja.dbpedia.org/data/ 葛飾北斎 .ttl](http://ja.dbpedia.org/data/葛飾北斎.ttl)」に 303 転送される。Turtle や RDF/XML のようなデータファイルを返答する場合は、SPARQL の DESCRIBE 文の結果を各形式で直接返すという実装になっている。

*8 <http://www.w3.org/2001/tag/group/track/issues/14>

*9 <http://www.w3.org/2001/tag/group/track/issues/57>

*10 <http://github.com/openlink/virtuoso-opensource>

*11 <http://github.com/fumi/dbpedia-vad-il18n>

SPARQL の DESCRIBE は、指定した IRI のリソースについての RDF グラフを返答するものであり、実際にどのようなデータが返ってくるかは SPARQL の仕様上規定されておらず、実装依存となっている。Virtuoso における DESCRIBE の実装は、指定した IRI を主語または目的語として含むトリプル群を返答するという汎用的な実装方法をとっている。DBpedia の RDF データは、事物の IRI を直接の主語としたトリプルだけで記述されているので、Virtuoso の汎用的な方法を問題なくそのまま利用している。

一方で、あるデータセットでは、事物の表現として、ブランクノードなどを含んだ深い構造の RDF データを用いているかもしれない。例えば vCard Ontology [Iannella 14] を用いて住所を記述している場合、図 5 のようなデータになる。この場合、汎用的な方法で事物の IRI を直接の主語としたトリプルだけを返してしまうと、vcard:hasAddress の先のブランクノードを主語としたトリプルが含まれないことになる。しかし RDF データとしては、vcard:hasAddress 以降も全部含んだトリプルを返したほうが、事物の表現として望ましいと考えられる。そうした場合に、DESCRIBE の実装として vcard:hasAddress の先のトリプルも含んで返すようにしてもよい。

```
@prefix : <http://example.org/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .

: 太郎 a vcard:Individual ;
vcard:hasAddress [ a vcard:Home;
vcard:postal-code "000-0000";
...
];
.....
```

図 5 vCard による住所記述例

3.4 原則 4：外部 IRI へのリンク

Web の特徴の一つであるリンクは、通常は HTML 表現内に含まれている。ユーザがある IRI にアクセスしたときに返ってくる HTML 表現に他の IRI へのリンクが含まれていると、ユーザはそのリンクを辿ることでさらに関連する HTML 表現などを読むことができる。Linked Data の四つ目の原則では、三つ目の原則である IRI にアクセスしたときに返ってくる RDF データに、外部のデータセットにある IRI へのリンクも含んでおくことで、ユーザが関連するデータを芋づる式に引っ張ることができるようになることを意図している。そのためにも、これまで説明してきたように IRI が RDF データについて参照解決可能になっている必要がある。

HTML におけるリンクと RDF におけるリンクの最大の違いは、HTML のリンクは単に文書から文書への

リンクでしかないが、RDF ではリンク自体もリソースであり、IRI で識別されているということである。DBpedia Japanese に存在する外部リンクは、DBpedia や DBpedia German といった他言語の DBpedia へのリンクと、日本語 Wikipedia オントロジー [玉川 13] へのリンクである。これらはすべて元来 Wikipedia のデータをもとにしているため、同一のリソースという関係を示す IRI である「<http://www.w3.org/2002/07/owl#sameAs>」を使用している。

実際に外部のデータセットへリンクをしようとしたときに、どこにリンクをするべきか悩むかもしれない。例えば研究データを Linked Data として公開するとき、まずは関連する専門的なデータセットにリンクすることが考えられるだろう。一方で、DBpedia Japanese のような汎用的なデータセットにもリンクすることで、異分野の意外なデータセットとつながる機会が増える。データが緩く間接的につながることで、公開したデータセット自体も価値が増大する可能性がある。これこそが Linked Data として公開することの意義であろう。

4. LOD クラウド

1 章で述べた LOD クラウドは、各データセットが他のどのデータセットとの間に RDF リンクを有しているかを調査して、図式化したものである。LOD クラウドに掲載されるようになるためには、一定の基準とプロセスが用意されている。2011 年の版までは、Open Knowledge が運営しているデータカタログサイトである the Datahub^{*12} に候補となるデータセットを所定の方式で登録しておくことで、LOD クラウドの作成者が基準を満たしているデータセットを lodcloud グループ^{*13} に追加するという行っていた。

2014 年の最新版 (図 2) では、lodcloud グループにあるデータセットに追加して、Billion Triple Challenge 2012 のデータセット^{*14}、public-lod@w3.org メーリングリストに投稿された計 56 万件の IRI を収集しており、それらを起点として、RDF リンクを順次たどることによって、対象データセットの収集を行っている [Schmachtenberg 14b]。

いずれにしろ、基準を満たしているデータセットの公開者が、the Datahub に所定の方式で登録しておくことが LOD クラウドに入るうえで重要である。そこで、LOD クラウドの基準とその手続きについて述べる。

4.1 LOD クラウド基準

LOD クラウドでは、Linked Data の原則に則ったデー

*12 <http://datahub.io>

*13 <http://datahub.io/organization/lodcloud>

*14 <http://km.aifb.kit.edu/projects/btc-2012/>

タセットかどうかを以下の基準で判定している。

- 参照解決可能な HTTP IRI でなければならない。
- よく使われているファイル形式の RDF データに、内容折衝があるいは別の方法で解決可能でなければならない。
- データセットは少なくとも 1 000 トリプル以上含む。
- すでに LOD クラウドにあるデータセットとの RDF リンクが 50 以上ある。
- RDF クローリングか、RDF ダンプ、あるいは SPARQL エンドポイントによって、データセット全体へアクセスできなければならない。

つまり、ここでは SPARQL エンドポイントや RDF ダンプだけ提供していて IRI が参照解決可能でないデータセットや、他との RDF リンクを含んでいないデータセットは対象外となっている。また、LOD クラウドは Web 上で実際につながっている Linked Data を図に示すことを目的としているため、各データセットがオープンライセンスで提供されているかどうかは求めていないことも注意が必要である。

4.2 the Datahub への登録と検証

LOD クラウドの基準を満たしているデータセットを次の LOD クラウドの候補にするには、まずデータセットについての情報を the Datahub に所定の形式で登録する必要がある。The Datahub は、誰もがデータセットについての情報を自由に登録して共有するためのサイトであり、Open Knowledge が CKAN*15 を用いて運営している。実際の登録方法については、Guidelines for Collecting Metadata on Linked Datasets in Data Hub*16 にまとめられている。

DBpedia Japanese は the Datahub に「http://datahub.io/dataset/dbpedia-ja」として登録されており、追加情報にデータセットのトリプル数(フィールド名 triples)や、外部リンク数(フィールド名 links:xxxx)などを記載されている(図6)。xxxx にくるのは他のデータセット名の the Datahub 上での名前であり、例えば DBpedia へのリンクは links:dbpedia というフィールド名になる。なお残念ながら、日本語 Wikipedia オントロジーについてはまだ登録されていないため、links: を付けずに参考情報として記載している。

The Datahub の基盤となっている CKAN の特徴として、データセットの各種例や実ファイルなどをリソースとして登録することができる。CKAN におけるリソースは Web における抽象化のリソースとは別概念であることに注意したい。LOD クラウド用の情報として、

*15 http://ckan.org
 *16 http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKANmetainformation



図6 the Datahub 上の DBpedia Japanese

CKAN 上のリソースに RDF の取得例や SPARQL エンドポイントへのリンク、データダンプへのリンクなどを登録することが求められている。DBpedia Japanese では N3 や RDF/XML の例や、SPARQL エンドポイントなどへリンクをしている。

データセットへのタグも、指定した形式で入力するように求められている。例えばデータセットに用いている既存語彙がある場合、format-foaf や format-geo といったタグを追加する。LOD クラウド上の分類もタグで指定する。DBpedia Japanese の場合は crossdomain を指

```
.....
@prefix : <http://ja.dbpedia.org/void.ttl#> .

<http://ja.dbpedia.org/void.ttl> a
  void:DatasetDescription ;
dcterms:title "A VoID Description of DBpedia
  Japanese"@en ;
dcterms:creator <http://fumi.me/foaf#me> .

:DBpediaJA a void:Dataset ;
dcterms:title "DBpedia Japanese"@en, "DBpedia
  Japanese"@ja ;
dcterms:description "DBpedia Japanese provides
  RDF datasets
  extracted from Japanese Wikipedia" ;
dcterms:creator <http://fumi.me/foaf#me> ;
dcterms:publisher :LODAC ;
dcterms:contributor :NII, :ROIS,
  :DBpediaCommunity ;
.....
```

図7 VoID 例

定している。その他、独自語彙は参照解決可能かどうかなど、データセットを把握するためのさまざまなタグが用意されている。

また、RDF データセットを記述するための語彙である VoID [Alexander 11] で、データセット自体について記述しておくことが推奨されている。DBpedia Japanese の VoID 「<http://ja.dbpedia.org/void.ttl>」の抜粋は図 7 である。

The Datahub に必要事項を記載したら、Data Hub LOD Validator^{*17} を使って項目の自動チェックを行うことが可能である。dataset の欄に the Datahub のデータセット名を指定することで、LOD クラウドに対する適合レベルと改善案を提示してくれる。

5. ま と め

本稿では DBpedia Japanese を例に、Linked Data が Web アーキテクチャのデータへの応用例であることを解説した。Linked Data という語が世に出てからすでに 10 年近く経つが、その間に着実に Linked Data の原則に基づくデータは増えてきている。日本語のデータにおいても、ここ数年でやっと事例が出てきているところであるが、まだまだ数は少ない。実際に LOD クラウドに入っているのは数えるほどである。本稿が少しでも日本語 Linked Data の増加に貢献できれば幸いである。

◇ 参 考 文 献 ◇

- [Alexander 11] Alexander, K. and Cyganiak, R., et al.: Describing linked datasets with the VoID vocabulary, *W3C Interest Group Note* (2011)
- [Berners-Lee 05] Berners-Lee, T., Fielding, R. and Masinter, L.: RFC3986: Uniform resource identifier (URI): Generic syntax, *IETF RFC 3986* (2005)
- [Berners-Lee 06] Berners-Lee, T.: Linked Data - Design Issues, <http://www.w3.org/DesignIssues/LinkedData.html> (2006)
- [Cyganiak 14] Cyganiak, R., Wood, D. and Lanthaler, M.: RDF 1.1 concepts and abstract syntax, *W3C Recommendation* (2014)
- [Dürst 05] Dürst, M. and Suignard, M.: Internationalized resource identifiers (IRIs), *IETF RFC 3987* (2005)

- [Fielding 99] Fielding, R., Gettys, J. and Mogul, J., et al.: Hypertext transfer protocol, HTTP/1.1, *IETF RFC 2616* (1999)
- [Hakala01] Hakala, J. and Walravens, H.: Using International standard book numbers as uniform resource names, *IETF RFC 3187* (2001)
- [Harris 13] Harris, S. and Seaborne, A.: SPARQL 1.1 query language, *W3C Recommendation* (2013)
- [Iannella 14] Iannella, R. and McKinney, J.: vCard ontology - For describing people and organizations, *W3C Interest Group Note* (2014)
- [Jacobs 04] Jacobs, I. and Walsh, N.: Architecture of the world wide web, *W3C Recommendation* (2004)
- [Kato 13] Kato, F., Takeda, H., Koide, S. and Ohmukai, I.: Building DBpedia Japanese and linked data cloud in Japanese, *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, 2014 年度人工知能学会全国大会 (第 28 回) (2014)
- [Lehmann 15] Lehmann, J., Isele, R. and Jakob, M. et al.: DBpedia - A Large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web Journal*, Vol. 6, pp. 167-195 (2015)
- [Sauerermann 08] Sauerermann, L. and Cyganiak, R.: Cool UR is for the semantic web, *W3C Interest Group Note* (2008)
- [Schmachtenberg 14a] Schmachtenberg, M., Bizer, C., Jentzsch, A. and Cyganiak, R.: *Linking Open Data Cloud Digram 2014* (2014)
- [Schmachtenberg 14b] Schmachtenberg, M., Bizer, C. and Paulheim, H.: Adoption of the linked data best practices in different topical domains, *The Semantic Web - ISWC 2014*, Vol. 8796 of Lecture Notes in Computer Science, pp. 245-260, Springer (2014)
- [玉川 13] 玉川 奨, 香川宏介, 森田武史, 山口高平: 日本語 Wikipedia オントロジーの構築と利用, 人工知能学会第 29 回セマンティックウェブとオントロジー研究会 (2013)

2015 年 6 月 4 日 受理

著 者 紹 介



加藤 文彦 (正会員)

2004 年慶應義塾大学大学院政策・メディア研究科修士課程修了。2004～07 年慶應義塾大学大学院政策・メディア研究科助手。2006～08 年 World Wide Web Consortium Systems Team。2007～09 年慶應義塾大学大学院政策・メディア研究科助教。2010 年未来技術研究所。2010 年～現在、情報・システム研究機構特任研究員。専門は Web 技術。LODAC や

DBpedia Japanese など、Linked Data やオープンデータに関する研究開発に従事。

*17 <http://validator.lod-cloud.net>