

# Linked Open Dataの技術と動向

武田 英明 ●国立情報学研究所 教授

**Linked Open Data (LOD) は文書ではなくデータのWeb。国内では2013年頃よりオープンデータの盛り上がりにもない関心を集める。**

## ■データのWeb

Webは1990年代初頭に学術情報の共有のために発明されたが、瞬く間に学術世界のみならず社会全体に普及し、われわれの情報流通の仕組みを大きく変化させた。このWebによる情報流通の革命と同じことが今、データの流通に起きつつある。それが、Linked DataあるいはLinked Open Data (LOD) である。

これまでのWebは、主に文書的情報を相互にリンクしてネットワークを作っている。いわば「文書のWeb」といえる。これに対して、LODは同様のネットワークをデータの間で作る。このため、LODは「データのWeb」と呼ばれる。

ただし、「文書のWeb」ではリンクには種類がなかったのに対して、「データのWeb」ではデータ間の関係を示すラベルをもつリンクになる。これにより、「このデータとこのデータは同一」「この項目の定義はこれ」といった関係を作ることができる。

LODはWebと同じようにグローバルに共有するデータ空間である。すなわち、データが世界のどこのだれが管理しているデータセット（データベース）に含まれているかを意識することなく、アクセスしたり、リンクしたりできる。これまであった、データセットとデータセットの間の障壁は存在しない。個々のデータセットはお互いにリ

ンクし合うことで、一つのグローバルなデータセットの一部となるわけである。

## ■LODの仕組み

実はLOD技術は、セマンティックWeb研究で作られた技術をデータの表現に利用したもので、多くの標準がすでに整備されている。セマンティックWebとはティム・バーナーズ＝リー (Tim Berners-Lee) 氏が提唱した、現行のWebより高度に知識を記述できるWebを作るというビジョンである。そのポイントはWebのグローバルな情報共有空間はそのままに、その上に、標準的なメタデータの記法であるRDF (Resource Description Framework) や、そのスキーマを記述する言語 (RDF SchemaやOWL) を用意することで、グローバルに知識を共有する仕組みである。

まず、RDFという言語ですべてを書く。RDFは、「主語」「述語」「目的語」に相当する3語の組み合わせ (3つ組) で、すべての情報を書く言語である。たとえば、「AさんはBさんを知っている」という関係は、「A knows B」という3つ組で表現される。このとき表現したい事物 (たとえば、前述の例ではAとB) には個別のURIを与える。ここでは個々の事物にURIを振ることで、URIが世界中でユニークなIDとして使えることが重要である。

さらにデータを記述する様式が決まっている場

1  
2  
3  
4  
5  
6

合、その様式をスキーマとして別途定義する。RDFを拡張したRDF Schema (RDFS) という言語では、データ記述の様式をクラスとプロパティの組み合わせで定義する。クラスはある事物や事象を示し、プロパティはクラスがもつ属性やクラス間の関係の型を示す。RDFSはRDFの拡張であるので、クラスやプロパティの定義もURIを与えて、参照可能にする。個別の事物や事象はこれらのクラスのインスタンスとして記述される。このようなスキーマ定義は、データセットごとに用意する必要はない。すでに他で定義されたスキーマがあるときは、それを使うことができる。

URIとRDFを使うことで、データがどこに含まれているといったことを気にせずに、データを相互につなげることができる。これは新しいデータの世界である。ティム・バーナーズ＝リーはこのLODを普及させるために4つの原則を提唱している。

1. 事物にURIを使って名前付けしよう
2. 名前の参照がHTTP URIでできるようにしよう
3. URIを参照したときに関連情報が手に入るようにしよう
4. 外部へのリンクも含めよう

この原則に基づくと、データセットは相互につながり、データセットのネットワークができる。

## ■ LODが作る世界

現在、上記の原則に基づいたLODが多数公開さ

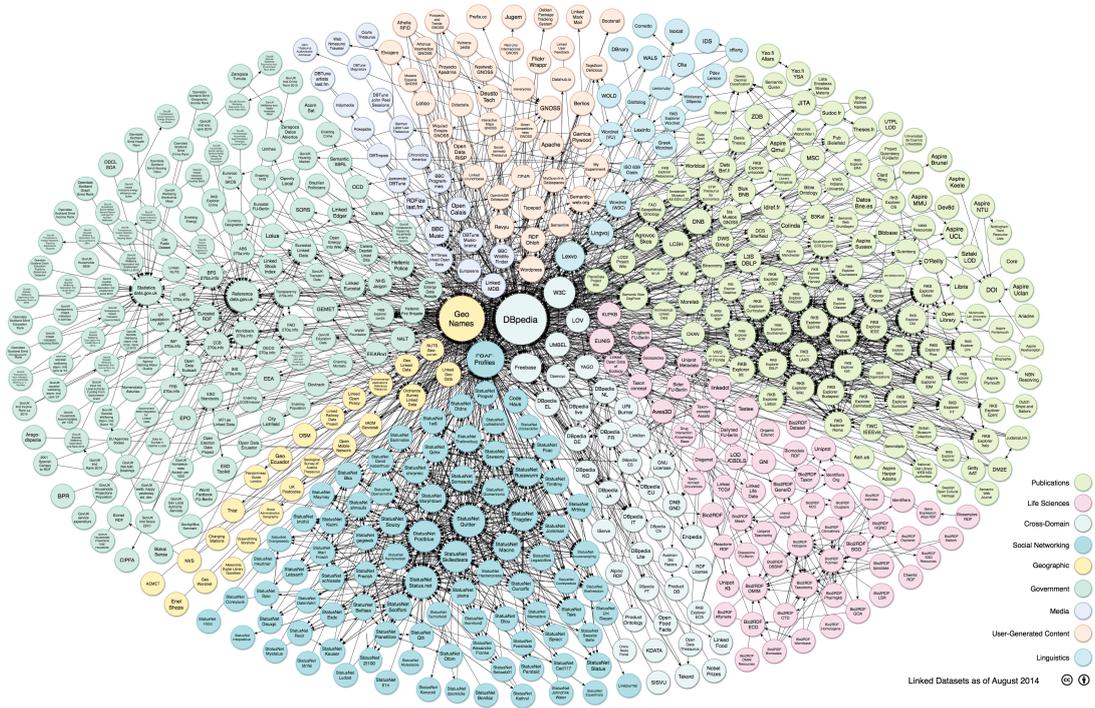
れている。資料5-2-3に示すのは、世界中のLODのネットワークを示す図で、これを「LOD Cloud」と呼んでいる。2014年8月時点の図で、500以上のデータセットが含まれている。丸が個別のデータセットを示し、データセットとデータセットを結んでいる線は、データセット内のデータ同士にリンクがあることを示している。

中心にあるのはDBpediaというもので、Wikipediaの情報をLOD化したデータセットである。Wikipediaは百科事典なので、さまざまな分野の項目が含まれている。そのため、他の多くのデータセットと結びつきやすい。これにより、DBpediaはこの図の中心にある。

資料5-2-3では、ドメインごとにデータセットが色分けされている。右上の部分は出版物に関連するデータセットで、図書館の書籍書誌データやジャーナル論文の書誌情報といったものが含まれている。右下はバイオサイエンスなどを含むライフサイエンスに関するデータセットである。左端は政府関係のデータセットである。ここには、アメリカの政府データポータルであるdata.govやイギリスのポータルdata.gov.ukのデータのうち、LODのものが含まれている。

こうした異なるドメインのデータが、DBpediaを中心にゆるくつながっていることがわかる。これらはみな前述のRDFで書かれているので、異なるデータセットのデータであってもあわせて使うことができる。

資料5-2-3 世界中のLODのネットワークを示す「LOD Cloud」



出典：Linking Open Data cloud diagram 2014<sup>1</sup>

## ■ LODの使い方

LODはこれまでのWebと同様に、検索サービスを通じてアクセスして使うこともできるし、リンクを順に辿ってみるといったブラウジングで使うこともできる。

しかし、LODの特徴を活かした使い方は、マッシュアップアプリケーションを通じての利用である。通常のマッシュアップアプリケーションでは、複数のWebサービスのWeb APIを利用してアプリケーションを構築する。それに対し、LODを使ったマッシュアップアプリケーションは、より容易に構築できる。LODはデータのアクセスの仕方が統一されているため、Web APIの場合のよう

に個別に対応を変える必要がない。また、記述形式もRDFで統一されているので、アプリケーションの中でも統一的に扱うことができる。さらに異なるデータセット間でも、リンクがあればそのまま使えるため、データ統合の手間を省くことができる。

たとえば、資料5-2-4は地図の上にさまざまな情報を配置する「Yokohama Art Navi」というアプリケーションである。一見、多くの地図アプリケーションと同じように見えるが、すべての情報が複数のLODのデータセット（博物館情報のLOD、アートイベントのLOD、地域のQ & AのLOD）から来ている。

資料5-2-4 マッシュアップアプリケーション「Yokohama Art Navi」



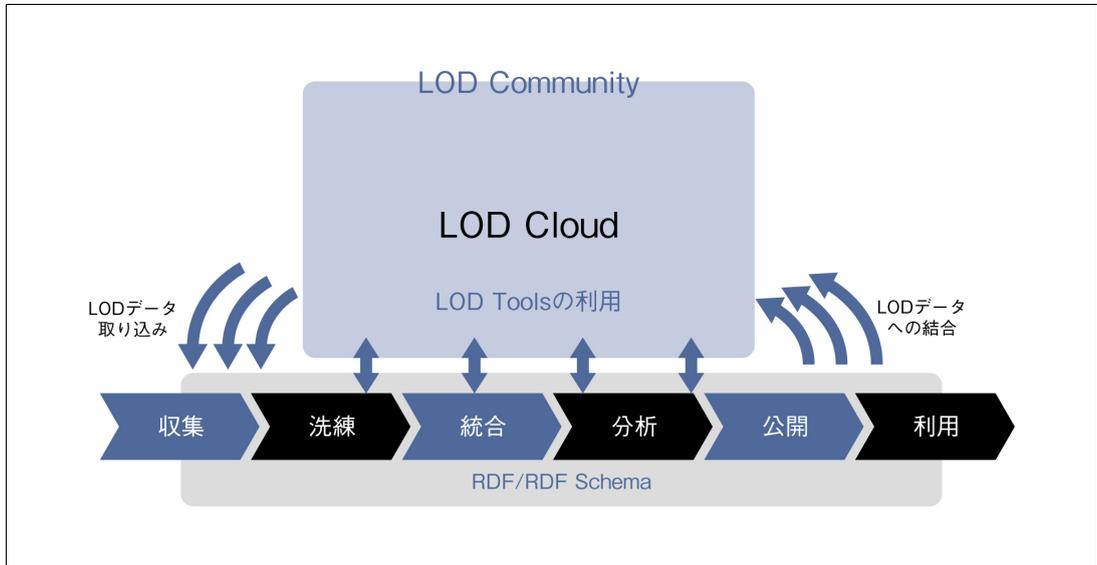
出典：Yokohama Art Spot<sup>2</sup>

■ LOD技術の展開

LOD技術の利用はデータの公開の場面に限定されているわけではない。データモデルをRDFのみとすることで、多種多様なデータを統合的に処理する、ETL (Extract/Transform/Load) の仕組みとして使うこともできる。

資料5-2-5に示すように、データの収集、洗練、

統合、分析、公開、利用といった一連のプロセスを、RDF処理だけで実現できる。その際、データやスキーマはそのまま外部から取り込むことができ、処理結果のほうもそのまま外部へ提供することもできる。また、さまざまなRDFに関わるツールやソフトウェアを、その間で使うことができる。



出典：筆者作成

## ■世界での動向

先に述べたように、LODの記述言語やアクセス言語は、World Wide Web Consortium (W3C)によって標準化されている。メタデータ記述言語RDFや、スキーマ記述言語RDF Schema、オントロジー記述言語OWL (Web Ontology Language)、RDFのデータベース (RDF StoreあるいはTriple Storeと呼ばれる) 問い合わせ言語SPARQLがこれにあたる。

共通に用いられる語彙やスキーマも多く提案され、主にデファクト標準として使われている。書誌的なメタデータ記述語彙であるDublin Core (これはDCMIによって標準化) や、人物と人物関係を記述する語彙FOAF (Friend-Of-A-Friend)、階層的な概念関係を記述するSKOS (Simple Knowledge Organization System)、統計データ記述語彙Data Cubeなどが有名である。

代表的なデータセットとして、Wikipediaの情報をLOD化したDBpediaや、アメリカ議会図書館の典拠データLODがある。また、アメリカと英

国のオープンデータをLOD化したものが多数作られている。そのほかのLODの利用例としては、遺伝子や化学物質データをLOD化したBio2RDFにおける統合的なデータ検索や、BBCの音楽番組や動物番組サイトでの利用などがある。

## ■国内での動向

国内では、2012年頃までは学術的関心が主であった。しかし、2013年頃よりオープンデータ活動が盛り上がり、それにともないオープンデータの次世代技術として広く関心を集めるようになった。

データセットとしては、日本語WikipediaをLOD化したDBpedia Japaneseや、国会図書館の典拠データや書誌データ、国立情報学研究所の論文検索サービスCiNiiのRDF化などが代表的なものである。また、統計センターでは都道府県・市区町村コードのLODも作られている。地方自治体関係では、横浜市芸術文化振興財団のヨコハマアートLODがある。

サービスとしては、ヨコハマアートLODを使っ

1

た「横浜MAPS」(iPhone アプリ) や、「さばえぶらり」などが提供されている。また、ETLに利用した例としては、横浜市金沢区の「かなざわ育なび」などがある。基盤的サービスとしては、Excel データを

2

RDF化する linkdata.org や LOD データのアクセスサービスなどを提供する LOD4ALL (富士通研究所) などがある。

3

---

1. Linking Open Data cloud diagram 2014, by Max Schacht-  
enberg, Christian Bizer, Anja Jentzsch and Richard  
Cyganiak. <http://lod-cloud.net/>  
2. <http://lod.ac/apps/yas/>

4

5

6



1996, 1997, 1998, 1999, 2000...

## [インターネット白書 ARCHIVES] ご利用上の注意

---

このファイルは、株式会社インプレスR&Dが1996年～2015年までに発行したインターネットの年鑑『インターネット白書』の誌面をPDF化し、「インターネット白書 ARCHIVES」として以下のウェブサイトで公開しているものです。

<http://IWParchives.jp/>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、データ、URL、名称など)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真・図の作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は掲載されていない場合があります。
- このファイルの内容を改変したり、商用目的として再利用したりすることはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用される際は、出典として媒体名および年号、該当ページ番号、発行元(株式会社インプレスR&D)などの情報をご明記ください。
- オリジナルの発行時点では、株式会社インプレスR&D(初期は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めました。すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

お問い合わせ先

株式会社インプレスR&D

✉ [iwp-info@impress.co.jp](mailto:iwp-info@impress.co.jp)