

ごぞんじですか? 第99回 Linked Open Data(LOD)とその活用事例

1. はじめに

Webはインターネット上の情報発信のための基盤技術となっています。

なぜWebが情報発信に多く用いられるようになったのか、その理由の一つに「情報を相互にリンクする」ことが挙げられます。Webページの情報同士は、ハイパーリンクを介して相互に渡り歩くことができます。

一方、これまでのハイパーリンクに加え、「構造化されたデータをWeb上で相互にリンクさせる」方法が現れました。これがLinked Dataです。

また、このLinked Dataという方法と、発信されている情報を自由に利用できるようにする「オープンデータ運動」が結びつくことで、「データを相互にリンクして共有する」Linked Open Data(LOD)として利活用ができるようになりました。

2. オープンデータ

オープンデータの法的・社会的・技術的な側面について議論しているオープンデータ・ハンドブック¹⁾では、Open Definition²⁾によるオープンデータの定義を以下のようにまとめています。

オープンデータとは、自由に使えて再利用もでき、かつ誰でも再配布できるようなデータのことだ。従うべき決まりは、せいぜい「作者のクレジットを残す」あるいは「同じ条件で配布する」程度である。

公的機関が保有する統計情報や地理情報、気象データなど、市民や民間企業でも活用が期待できる情報がWeb上には多々あります。しかし、日本の著作権法では、複製や転載については著作権者

の許諾を得る必要があるなど、「再利用」という観点からは制限があるといえます。

そこで、公開する情報について、あらかじめ一定の条件で利用を認めておけば、利用者は許諾を得る必要はありません。

オープンデータの代用的なライセンスにはクリエイティブ・コモンズ・ライセンス³⁾があり、このうちCC0(権利放棄)、CC BY(クレジットを表示し複製・再配布・改変可)、CC BY-SA(クレジットを表示し同一条件で再配布可。改変は不可)が用いられます。これらの条件を公開時に示すことで、著作物をこの範囲で自由に使うことができます。

3. LODとは

Linked Dataについては、Webの提唱者であるTim Berners-Leeが以下の4つの基本原則を示しています⁴⁾。特に、誰でも自由に利用できるオープンデータとして公開されたものは、LODと呼ばれます。

1. あらゆるものを識別するのにURI(Uniform Resource Identifier)を使いましょう。
2. HTTP URIを使うことで、アクセスしたときに参照解決可能にしましょう。
3. ユーザやエージェントがURIにアクセスしたら、RDFやSPARQL⁵⁾といったWeb標準の手段で役に立つ情報を提供しましょう。
4. (3.の情報に)関係する他のURIへのリンクを含みましょう。そうすることでさらに多くのものを発見することができます。

まず、インターネット上にあるもの(リソース)を識別できるようにすることが必要です(基本原

則1)。URIはWeb上でリソースの場所を表すURL(Uniform Resource Locator)の考え方を拡張したもので、識別子(Identifier)としての役割を持つものです。

URIで示された場所にWebの標準プロトコル(HTTP)でアクセスすることで、何らかの情報が返ってくることで、またこの時に後述するRDFなどの標準的な手段が利用できることを求めているのが基本原則の2と3です。

また、返ってきた情報に他のリソースへのリンクが含まれていれば、利用者はリンクをたどってより多くの情報を得ることができます。これが基本原則の4です。

すでに多くの情報資源がLODとして公開されています。一般的な事物については、Wikipediaから抽出した情報で作成されているDBpedia⁶⁾があります。図書館の世界では、国立国会図書館や米国議会図書館、英国図書館などの典拠データがLODで公開されています⁷⁾。

4. LODの要素技術とその例

リソースを記述するための枠組みとして、RDF(Resource Description Framework)が用意されています。RDFでは、リソースを主語(subject)、述語(predicate)、目的語(object)の3つ組み(トリプル)により表現します。

情報検索の世界で用いられるシソーラスも、この3つ組みで表すことができます。例えば、国立国会図書館件名標目表のうち「専門図書館」という主題については、日本語では以下のように表すことができます。

- 「専門図書館」の上位の概念(BT)には「図書館」があります。
- 「専門図書館」の下位の概念(NT)には「農学図書館」があります。
- 「専門図書館」はNDC9版の分類番号018と関連があります。

国立国会図書館件名標目表をLODとして提供

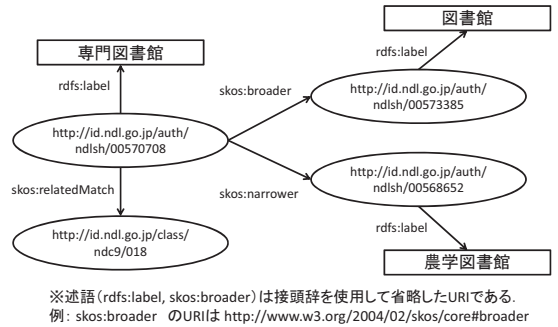


図1 RDFによる主題の図示

している「国立国会図書館典拠データ検索・提供サービス(Web NDL Authorities)」で同じ「専門図書館」という主題をRDFで表したものの一部が図1です。主語、述語、目的語のいずれもURI(<http://id.ndl.go.jp/auth/ndlsh/00570708>など)で表されているほか、人間が理解できるよう、主語と目的語それぞれに対する説明(rdfs:label)として「専門図書館」などの日本語が加えられています。

また、述語のうちskos:broaderは、シソーラスのような概念体系の基本構造や内容を表現するための語彙であるSKOS(Simple Knowledge Organization System)に含まれるもので、「より広義な概念」を表します。その意味は、URI(<http://www.w3.org/2004/02/skos/core#broader>)を参照することで知ることができます。

主語や目的語を表現するための語彙は、データ同士を相互にリンクしやすくするためにすでに存在するものを再利用することが多いです。タイトル、作成者、日付、主題などの汎用的な語彙にはDublin Core Metadata Initiative(DCMI)メタデータ語彙、人物の名前や相互の関係にはFriend-of-a-Friend(FOAF)語彙など、すでに多くの語彙が利用されています。

DCMIメタデータ語彙を使って、DBpedia上の「専門図書館」という語とWeb NDL Authoritiesの「専門図書館」という主題をリンクしたものが図2です。このうち<http://purl.org/dc/elements/1.1/subject>は「主題」を、<http://www.w3.org/2004/02/skos/core#relatedMatch>は「語彙同士の関連」を表す語彙です。図2では説明(ラベル)

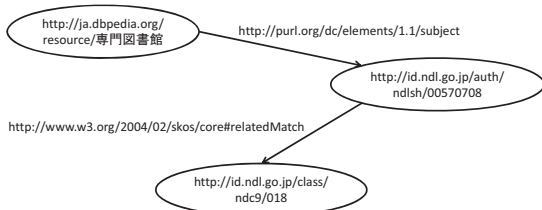


図2 DBpediaとWeb NDLのリンク

は省略しています。

- (DBpediaの)専門図書館の主題は(Web NDL Authoritiesの)「専門図書館」です。
- (Web NDL Authoritiesの)「専門図書館」はNDC 9版の分類番号018に関連があります。

この2者のリンクを通じて、「(DBpediaの)専門図書館はNDC 9版の分類番号018に関連があります」と表すことができます。

このように、「Web上にあるリソース」をURIで識別可能にすることで、異なる情報資源同士でもお互いにリンクし繋がってゆくことができます。

5. 農学分野での例

FAO(国際連合食料農業機関)が構築、提供している農学分野の文献データベースAGRIS⁸⁾は、LODを活用することで文献情報以外の様々な情報資源とのリンクを可能にしています⁹⁾。

図3はAGRISを稲の学名である「Oryza sativa」



図3 AGRISの検索結果詳細表示画面

で検索した結果の一例です。他の文献データベースとの大きな違いは、他の情報資源とのリンクといえます。図3の右側、「Related information in other data sources」には、Googleの検索結果を基にした関連文献のほか、World Bankが提供するCereal yield(穀物収量)の国別統計、またDBpediaによる用語解説などが表示されています。

他の情報資源とのリンクには、AGRISに使用されているシソーラス、AGROVOC¹⁰⁾が大きな役割を果たしています。検索結果(図3)から「RDF」と書かれたボタンをクリックするとRDF形式のデータが表示されます(図4)。このうち、「source」と「subject」がURIで表されていることがわかります。sourceはこの書誌情報を入力したインプットセンターを表します。subjectはAGROVOC上の統制語です。このうち「rice」にあたるhttp://aims.fao.org/aos/agrovoc/c_6599.htmlを表示したものが図5です。米国議会図書館件名標目表やDBpedia上の情報とskos:exactMatch(同等の概念)という関係でリンクされています。

このように、AGRISではAGROVOCを介して米国議会図書館件名標目表など既存の情報源とのリンク先の情報を参照することで、図3のような

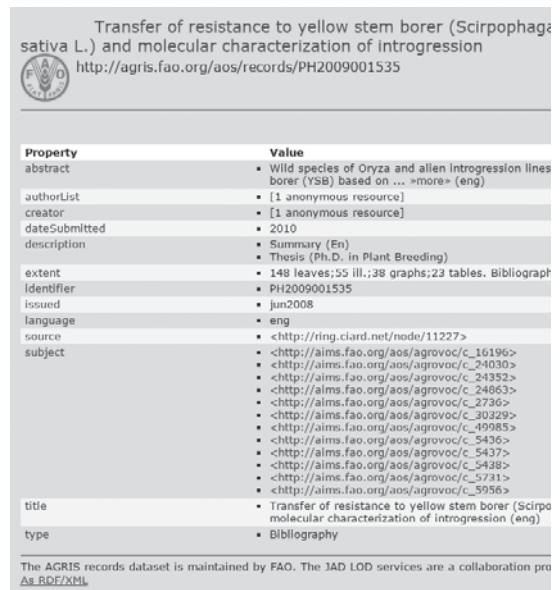


図4 RDF形式での表示

Property	Value
rdftype	skos:Concept AGROVOCの上位語、下位語
skos:broader	広義の概念 http://aims.fao.org/aos/agrovoc/c_1474 http://aims.fao.org/aos/agrovoc/c_1474
skos:narrower	狭義の概念 http://aims.fao.org/aos/agrovoc/c_330606 http://aims.fao.org/aos/agrovoc/c_330653
skos:exactMatch	同等の概念 http://www.eionet.europa.eu/gemet/concept/7214 http://lod.nal.usda.gov/nalt/56293 http://eurovoc.europa.eu/ 米国家農学図書館作成のシソーラス http://d-nb.info/ http://id.loc.gov/authorities/sh85113862#concept http://www.caas.gov/ 米国家議会図書館件名標目表 http://zbw.eu/stw/descriptor/14095-0
skos:closeMatch	http://dbpedia.org/resource/Rice
isTerms:created	2011-11-20T20:05:00 DBpedia

図5 AGROVOCとそのリンク先

関連情報の表示を行っています。これらはURIにより参照していますので、利用者は直接そのデータを参照し、必要な情報を得ることができます。また、この関係を利用して、逆に他の情報資源からAGRISの書誌情報を参照することもできます。

6. まとめ

LODという方法により、データ同士をリンクした新たな活用が可能となりました。

今後、LODにより他とのリンクが可能な情報資源がますます増えてゆくことが想定されます。一方で、これまで図書館で構築され書誌情報の作成や情報検索で使われてきた件名表目標やシソーラスが、これらの情報資源同士をつなぐという新たな用途でも活用されるようになりました。LODは図書館においても関連のある技術といえます。

AGRISの場合は、書誌情報のようにすでに構造化されたデータが存在したこと、特にAGROVOCのような他の情報資源と照合しその関係を把握することができる情報源を備えていたことなどから、LODへの対応も比較的容易に行えたと考えられます。

同様に、これまでWeb上に蓄積された情報がWebページだけではなくデータを単位として相互にリンクされることで、情報資源のさらなる活用が期待できます。

国際農林水産業研究センター
林 賢紀(はやし たかのり)

参考文献

- 1) Open Knowledge Foundation. “オープンデータ・ハンドブック”. <http://opendatahandbook.org/guide/ja/>. (参照2015-11-20)
- 2) Open Knowledge Foundation. “オープンの定義”. <http://opendefinition.org/od/1.1/ja/>. (参照2015-11-20)
- 3) クリエイティブ・コモンズ. “クリエイティブ・コモンズ・ライセンス”. <http://creativecommons.jp/>. (参照2015-11-20)
- 4) Tim Berners-Lee: "Linked Data". <http://www.w3.org/DesignIssues/LinkedData.html>. 2006. (参照2015-11-20) なお、本項での日本語訳は参考文献の5)を参考とした。
- 5) SPARQLはRDFを検索、操作するための言語である。詳細は以下を参照されたい。加藤文彦[ほか]. “オープンデータ時代の標準Web API SPARQL”. インプレスR&D. 2015.
- 6) DBpedia Community. “DBpedia Japanese”. <http://ja.dbpedia.org/>. (参照2015-11-20)
- 7) 橋詰秋子. “Linked Open Data—欧米国立図書館の動向と国立国会図書館の取組み”. NDL書誌情報ニュースレター. 2015年2号(通号33号). http://www.ndl.go.jp/jp/data/bib_newsletter/2015_2/article_01.html. (参照2015-11-20)
- 8) FAO. “AGRIS”. <http://agris.fao.org/>. (参照2015-11-20)
- 9) 林賢紀. “E1523 - 農業関連の文献データベースAGRISとそのリニューアルについて”. カレントアウェアネス-E, No.252, 2014, <http://current.ndl.go.jp/e1523>. (参照2015-11-20)
- 10) FAO. “AGROVOC Multilingual agricultural thesaurus”. <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>. (参照2015-11-20)