

4

# 出版物に関するメタデータと 国際書誌コントロール

## —国立国会図書館における LOD の取り組み—

応  
般

橋詰秋子（国立国会図書館 電子情報部） 福山樹里（国立国会図書館 電子情報部）

### 国立国会図書館の LOD の取り組み

#### ● LOD 提供の背景

国立国会図書館（National Diet Library：NDL）は、我が国唯一の国立図書館であり、国会の国政審議を補佐する機関として立法府に属している。1948年の設立以来、国立国会図書館法に定められた納本制度に基づき日本の文化的財産・知的資源として国内出版物を収集し、それらを蓄積し後世に伝え（保存し）、広く利用に供するという役割を担ってきた。蔵書数は国内最大で、2015年3月末時点で4,100万点を超える。近年では蔵書のデジタル化や各種データベースの提供など、インターネットを通じた電子情報サービスに力を入れている。

NDLは2010年以降、電子情報サービスの一環として、APIを通じて、蔵書目録のデータを初めとする各種データをRDF形式で提供してきた。現時点では、NDLが提供するリンクトデータにオープンライセンスは適用されておらず、狭義にはオープンデータとはいえないものもあるが、本稿ではそれらも含めて「NDLのリンクト・オープン・データ（LOD）」

と呼ぶことにする。なお、本稿の意見にかかる部分は筆者らの見解であって、所属機関を代表するものではないことをあらかじめお断りしておく。

2016年4月現在、NDLは、表-1に挙げた4種類のデータをLODとして公開している。

#### ● 課題の把握と改善の取り組み

最近、オープンデータのムーブメントの盛り上がりが目覚ましいが、NDLのLODはこの動きの中で進められたものではなかったため、認知度が低く活用事例が少ないといった課題があった。そのため、2014年にNDLのLODの現状やニーズを探る調査を行い、その結果を踏まえて課題と改善策を整理した。

NDLのLODの課題は、大きく分けて、(1)認知度の向上、(2)利活用の促進、(3)提供するLODの充実、の3つがあると考えられた。現在は、着手可能なものから改善に取り組んでいる。

##### (1) 認知度の向上

2014年9月、NDLのLODをまとめて紹介するWebページ<sup>1)</sup>を作成した。それまで、NDLのLODは各電子情報サービスのWebサイトに説明ページがあるだけで分かりにくかった。そこで、各説明ペ

種類	概要	データを提供している電子情報サービス	公開
書誌データ	出版物のメタデータ（詳細は本文参照）	国立国会図書館サーチ <a href="http://iss.ndl.go.jp/">http://iss.ndl.go.jp/</a>	2010.8
典拠データ	複数の名前を持つ著者や同義語を同定したり同名異人を識別したりするための情報（詳細は本文参照）	国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities） <a href="http://id.ndl.go.jp/auth/ndla/">http://id.ndl.go.jp/auth/ndla/</a>	2011.7
震災関連データ	東日本大震災に関する資料や画像、動画のメタデータ。タイトル、作成者、保存日、撮影場所、要約・抄録などを含む	国立国会図書館東日本大震災アーカイブ <a href="http://kn.ndl.go.jp/">http://kn.ndl.go.jp/</a>	2012.11
図書館コード	日本の図書館を対象とした識別子。図書館の種類、URL、位置情報、電話番号などの組織情報を含む	オープンデータセット <a href="http://www.ndl.go.jp/jp/aboutus/standards/opendataset.html">http://www.ndl.go.jp/jp/aboutus/standards/opendataset.html</a>	2015.4

表-1 NDLが提供しているLOD

ージへの入口として、NDLのLODの概要、取得方法、利用条件などが1カ所で分かるページを新設した。ほかにも、オープンデータ分野での認知度を上げるために、LODチャレンジやアーバンデータチャレンジなどのデータ利活用コンテストに積極的に参加し、関連イベントでのデータ紹介などを行っている。

## (2) 利活用の促進

アイデアソン、ハッカソンなどのイベントを通じて、利活用の促進を図っている。2015年2月に「国立国会図書館のWebページを使い尽くそうアイデアソン」を、同年11月に「国立国会図書館のデータを使い尽くそうハッカソン」を開催した。イベントの参加者によって、スマートフォン用のアプリケーションやソーシャルゲームなど、従来にない新しいユースケースが生まれ始めている。

## (3) 提供するLODの充実

以前から提供してきたLODに加え、ニーズの高かった各種コード類のリンクトデータ化を試みている。たとえば、これまではCSV形式で提供してきたISO規格の図書館コード「図書館および関連組織のための国際標準識別子 (ISIL)」をリンクトデータ化し、2015年4月から試行提供を開始した。また、日本の図書館の標準的な資料分類法である「日本十進分類法」のリンクトデータ化を目指して、2015年度から、日本図書館協会とNDLとで共同研究を実施している。

## 国立国会図書館のLODの紹介

NDLが提供するLODのうち、利用されること多い「書誌データ」と「典拠データ」を紹介する。

### ● 書誌データのLOD

#### 書誌データとは

書誌データとは、簡単にいえば、出版物のメタデータのことであり、蔵書目録 (NDL-OPAC) 構築の一環で作成されている。書誌データが対象とする出版物は、図書だけでなく、雑誌、記事、論文、新聞、デジタルコンテンツ (NDLがデジタル化

した資料、Webページ) などである。メタデータには、タイトル、著者、出版者、件名 (テーマ、地名)、分類、ISBNなどの識別子、著作権関係の情報、所蔵館などの情報が含まれている。

書誌データに含まれる多様な情報をWeb上のほかの情報と組み合わせて使うには、識別子 (ID) の活用が重要となる。識別子には、国際的なルールに従って付与されるものと提供機関が独自に定めたルールで付与されるものがある。NDLの場合、前者の例には、図書に付与される「ISBN (国際標準図書番号)」や論文などに付与される「DOI (デジタルオブジェクト識別子)」が、後者の例には、NDLが作成・提供する書誌データに付与される「NDLBibID (国立国会図書館書誌ID)」やデジタル化資料やWebアーカイブ資料等に付与される「NDLJP (NDLが付与する永続的識別子)」がある。なお、書誌データをより効果的に利用できるよう、これらの識別子をリスト化した「書誌IDリスト」を、自由に二次利用が可能なオープンデータセットとして公開している。

#### 書誌データのLODの提供

書誌データは、「国立国会図書館サーチ (NDLサーチ)」(図-1) のAPIで、RDF/XML、XMLなどの形式で取得することができる。APIは、Amazonの子会社が開発したOpenSearchやOAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) など、出版物のメタデータの検索やハーベストに適した各種プロトコルに対応している。このAPIのアクセス数は、年間約1,124万件である。データの記述は、メタデータ記述の国際標準である



図-1 国立国会図書館サーチ

```

</dctypes:Identifier>
<dctypes:Identifier rdf:datatype="http://ndt.go.jp/dctypes/ISBN">4-06-261654-8</dctypes:Identifier>
<rdftype:seeAlso rdf:resource="http://ndt.go.jp/jpno/960654247">
<rdftype:seeAlso rdf:resource="http://iss.ndt.go.jp/ISBN/49626165487">
<dctypes:title>めぐりめぐる月</dctypes:title>
<rdf:Description>
  <rdf:value>めぐりめぐる月</rdf:value>
  <dctypes:transcription>メグリメグルノヅキ</dctypes:transcription>
</rdf:Description>
</dctypes:Alternative>
<rdf:Description>
  <rdf:value>Walk two moons</rdf:value>
</rdf:Description>
</dctypes:Alternative>
<dctypes:SeriesTitle>シリーズタイトル</dctypes:SeriesTitle>
<rdf:Description>
  <rdf:value>ユスヤレクション</rdf:value>
</rdf:Description>
</dctypes:SeriesTitle>

```

図-2 書誌データに含まれるタイトル関連の情報（NDL 書誌データの抜粋）



図-3 Web NDL Authorities

Dublin Core をもとに、NDL 独自の拡張を加えた「国立国会図書館ダブリンコアメタデータ記述 (DC-NDL)」に従っている。利用にあたっては、営利目的の場合は利用申請が必要だが、非営利目的の場合は手続き不要となっている（ただし、非営利目的の場合でも、継続的なアクセスを行う場合には利用申請をお願いしている）。

図書館が作成する書誌データは、オンライン書店のものとは比べると、精緻に構造化されている。たとえば、一口に“タイトル”といっても、図書館の書誌データには、本タイトル、本タイトルのよみ、別タイトル（別言語）、シリーズタイトル、シリーズタイトルの読み、といった情報が含まれており、それらが関連付けられる形で構造化されている（図-2）。

NDL は納本制度に基づく国内出版物の網羅的な収集を方針としているため、その蔵書を対象とした書誌データの LOD は、網羅性の高い日本の出版物のデータとして活用できる。この LOD を使えば、どんな出版物が存在しているか調べられるだけでなく、日本の出版物の傾向を分析することも可能だろう。

## ● 典拠データの LOD

### 典拠データとは

典拠データとは、図書館の分野で伝統的に構築されてきた「統制語彙 (controlled vocabulary)」である。統制語彙とは、検索システムのインデックスとして使用する語を限定し、その意味範囲や使用法

を規定した（コントロールされた）語彙である。蔵書目録において、同義語や同形異義語、言葉の意味のあいまいさなどによって生じる検索もれやノイズを回避することを目的に作成されている。複数の名前（筆名、旧姓など）を持つ著者（例：「夏目漱石」「夏目金之助」）や、資料の内容（テーマ）を示す語句について同義語や上位語・下位語・関連語などの情報を扱っている。

著者の場合、たとえば歌舞伎役者は襲名によって名前が変わっていくが、典拠データには「松本幸四郎 9 世」は「市川染五郎 6 世」と同一人物である、という関連を示す情報が含まれる。逆に、「鈴木一郎」という一般的な名前に対しては、同名異人の著者を識別する情報を「鈴木、一郎、1925- 牧師」「鈴木、一郎、1925-」といった形で保持している。さらに、資料の内容（テーマ）を表す語句（図書館分野では「件名」と呼ばれるもの）の場合、たとえば、「震災」には「地震災害」「Earthquakes」が同義語としてあり、上位語に「自然災害」、下位語に「地震火災」、関連語に「地震」「地震保険」「震災予防」があるといった情報が含まれている。

### 典拠データの LOD の提供

NDL では、典拠データの LOD を「国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities)」を通じて提供している（図-3）。RDF/XML, RDF/Turtle, JSON で取得できる SPARQL エンドポイントを提供している。また、資料の内容（テーマ）を表す語句を集めた「国立国会図



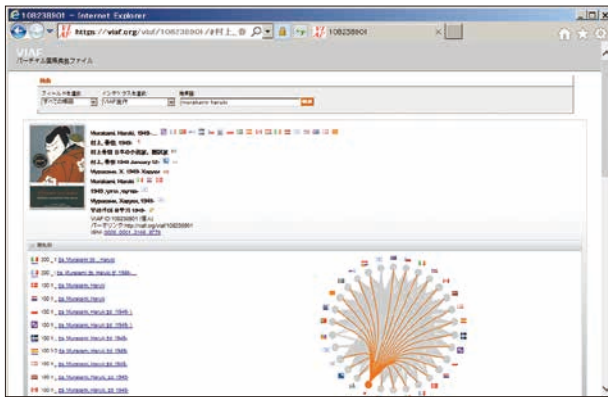


図-6 バーチャル国際典拠ファイル (VIAF)

最近になり、これらの活動と親和性が高い技術としてリンクトデータが取り入れられるようになった。

たとえば、英国図書館は、「Linked Open BNB」という名前を付けて、2010年から英国全国書誌 (BNB) に含まれる書誌データ (図書・雑誌) 280 万件を LOD で提供している。また、ドイツ国立図書館は、2010年から蔵書目録の典拠データをリンクトデータ化しており、2012年にはドイツ語圏の統合典拠ファイルの LOD へと拡張させている。典拠データは国際的な連携も進んでおり、「バーチャル国際典拠ファイル (VIAF)」<sup>3)</sup> という Web サービスでは、各国の図書館が作成した典拠データを集め、個人や団体といった同一の実体に対する各国の典拠データを同定し相互にリンクさせる形で、世界規模の仮想的な統合典拠ファイルを実現させている。VIAF には NDL も参加しており、前述の Web NDL Authorities の典拠データには VIAF へのリンクが含まれている。そのため、VIAF を介することで、図-6 のように日本の典拠データから諸外国の典拠データにリンクしている。

英国図書館の事例は別の形式で蓄積されていた既存データをリンクトデータへ変換するものであったが、目録のデータを作成する段階からリンクトデータを採用するアプローチも存在する。米国では、米国議会図書館 (LC) が積極的に LOD に取り組んでいるが、その中心は各種コード類と典拠データの LOD 化である。「LC Linked Data Service」では、米国議会図書館件名標目表 (LCSH)、LC 名称

典拠ファイルなど約 40 種類を LOD で提供している。この LC の事例には、その背景に、LC が中心となり進めている「BIBFRAME」と名付けられたプロジェクトがある。図書館の分野では、1960年代に「MARC」と呼ばれる専用のデータフォーマットが開発され、目録のデータに用いられてきたが、近年の情報技術の進展に十分に適用できずレガシーとなっていた。BIBFRAME は、MARC に代わる新しいデータフォーマットを開発するもので、そのデータモデルの基盤にリンクトデータが採用されている。LC が取り組んでいる各種コード類や典拠データの LOD は、目録で記述されるさまざまなリソース (分類や件名) に URI を与えるものであり、言い換えれば BIBFRAME/Linked Data を使った目録作成に必要な語彙を開発しているといえる。

NDL が行っている LOD の取り組みも、国際書誌コントロールの流れを受けて進められている。現在、この流れは、BIBFRAME のような目録の作成時点でのリンクトデータの採用に向かって動いていると考えられるが、この方向性は NDL も共有している。NDL では、2013年に、今後の書誌データ作成の方向性を示す文書「国立国会図書館の書誌データの作成・提供の新展開 (2013)」を策定し、この中で MARC の限界とリンクトデータへの対応など Web に適した書誌フレームワーク構築の必要性を指摘した。

## 将来に向けて

筆者らは、国際的な潮流となっている図書館による LOD の取り組みの裏には、さまざまなプレーヤが存在する Web 世界において、情報提供機関としての図書館の存在感を高めたい、という図書館員の意識があると推察している。今後、“データの Web”の中で図書館が提供するデータが有用な地位を獲得するためには、ニュースサイトや Wikipedia など、図書館以外の分野へ活用を広げていく必要があるだろう。そのためには、図書館のデータを、Web 世界に適した形で提供することが大切となる。

NDLのLODは、改善途上であり、まだ十分に利活用しやすいものとはなっていない。今後もできることから改善したいと考えている。そう遠くない将来に、NDLが提供する各種のLODが、図書館の枠を越えて広く使われるLODとなることを期待している。

### 参考文献

- 1) 国立国会図書館：使う・つなげる：国立国会図書館のLinked Open Data (LOD) とは, <http://www.ndl.go.jp/jp/aboutus/standards/lod.html>
- 2) Linked Web NDL Authorities, <http://indigo-lab.github.io/ndlna/>
- 3) バーチャル国際典拠ファイル, <http://viaf.org>  
(2016年3月24日受付)

橋詰秋子 ■ [hasizume@ndl.go.jp](mailto:hasizume@ndl.go.jp)

2001年、国立国会図書館入館。図書館協力部、関西館、国際子ども図書館等を経て、2014年から電子情報部電子情報流通課標準化推進係長としてLinked Open Dataやメタデータ、データ利活用の推進を担当

福山樹里 ■ [ju-fukuy@ndl.go.jp](mailto:ju-fukuy@ndl.go.jp)

2010年、国立国会図書館入館。主題情報部（現、利用者サービス部）科学技術・経済課科学技術係を経て、2013年から電子情報部電子情報流通課標準化推進係にてLinked Open Data、メタデータ、データ利活用の推進を担当

