

教科書 Linked Open Data (LOD) の構築と公開

江草 由佳*, 高久 雅生**

教科書 LOD は、国立教育政策研究所教育図書館や教科書研究センター附属教科書図書館が長年かけて組織化してきた書誌情報をまとめて LOD 化したものである。1992 年施行の学習指導要領以降の検定教科書を対象として、書誌事項と教科等の関連情報を LOD 化し、2018 年 3 月現在、7,257 タイトルの教科書情報、RDF データとして 157,297 トリプルを公開している。本稿では教科書 LOD の開発および公開を通して得た知見を紹介する。

キーワード：教科書, Linked Open Data, 書誌情報, Uniform Resource Identifier (URI)

1. はじめに

近年、構造化されたデータを共有するための枠組みとして Linked Open Data (LOD) が注目を集めている¹⁾。LOD は、セマンティックウェブを指向したデータ共有の枠組みで、さまざまな領域のデータをその記述の粒度を問わずに扱え、自由に活用できる特徴を持つ²⁾⁴⁾。図書館等が所蔵する書籍や論文などに対して組織化してきた書誌情報や、分類や件名等の主題情報、著者等の典拠情報を LOD 化して、誰もが利用できるように共有することにより、ウェブを始めとするさまざまな領域での多様な活用が期待でき、世界中で多くの情報機関が取り組むようになってきている⁵⁾⁶⁾⁷⁾。

著者らは、日本の現行の検定制度における教科書の書誌データおよび教科や学習指導要領などの教科書に関連する情報を LOD 化し、「教科書 LOD」として 2017 年 1 月から、「<https://w3id.org/jp-textbook/>」で公開を始めた。この教科書 LOD は、国立教育政策研究所教育図書館（以降、教育図書館と呼ぶ）や教科書研究センター附属教科書図書館（以降、教科書図書館と呼ぶ）が長年かけて組織化してきた書誌情報をまとめて LOD 化したものである。1992 年施行の学習指導要領以降の検定教科書を対象として、書誌事項と教科等の関連情報を LOD 化し、2018 年 3 月現在、7,257 タイトルの教科書情報、RDF データとして 157,297 トリプルを公開している。LOD データセットのコンテストである LOD チャレンジ 2016 に応募し、テーマ賞「教育 LOD 賞」を受賞した⁸⁾。本稿では教科書 LOD の開発および公開を通して得た知見を紹介する。

教科書 LOD とは、教科書を URI (Uniform Resource

Identifier)⁹⁾により一意に示し、当該教科書の書誌情報や教科などの基本的な情報を LOD として提供するものである。教科書 LOD の最大の特徴は、教科書の表紙や奥付に必ず記載されている、学校種別、教科書記号、教科書番号、検定年の 4 つからなる URI を使うことで、当該教科書を機械的に一意に示すことができるようになった点である（学校種別、教科書記号、教科書番号、検定年の詳細は後述する）。手元に教科書の現物があれば、この URI を使って教科書を指し示せるようになった。また、LOD 技術により、この URI を使って書名や出版者などの書誌情報に加えて、教科名や学年などの教科書特有の情報、さらには、学習指導要領や教科、科目までさまざまな情報を得られる。

本稿では、教科書 LOD の概要を掴んでもらうために、まず、教科書 LOD の閲覧・ダウンロード用ウェブサイトの利用イメージを説明し、その後、URI 設計などのデータの内容について説明する。

2. 教科書とは

現代の日本における初等中等教育は、文部科学省が作成した学習指導要領¹⁰⁾に基づいて、全国すべての学校が同じ教育課程の基準に基づく内容で、児童生徒が学習している。実際の教育にあたっては、それぞれの教科用に供される教科用図書（以下、教科書と呼ぶ）が重要なツールとなり、具体的な学習内容を伝えている¹¹⁾。

教科書は、検定制度に基づき、多大なコストをかけて制作される。検定制度のもとでは、教科書の企画から編纂、専門家による検定と合否判定、教科書の選定、採用まで、約 5 年以上の歳月を費やし、多くの専門家がかかわって教科書が制作されている。このようにして作成された教科書は、義務教育においては無償提供されている。平成 29 年度に使用された教科書のタイトル数は 897 種類で、上下巻などを別と数えると 1,461 種類あり、日本全国で約 1 億 2400 万冊が使われた¹²⁾。教科書に基づく学習や教科指導は、日本の教育力の基盤になっており、教育内容の標準化と高い学力水準¹³⁾を支えている。

図書館資料としての教科書は、同じ時代に教育を受けた

* えぐさ ゆか 国立教育政策研究所 研究企画開発部教育研究情報推進室

** たかく まさお 筑波大学 図書館情報メディア系
〒100-8951 東京都千代田区霞が関 3 丁目 2 番 2 号 中央合同庁舎 7 号館 6 階 6C
E-Mail: yuka@nier.go.jp (原稿受領 2018.5.2)

世代が共通して持つ世代記憶の一種となっており、図書館におけるレファレンス利用も多く、利用者からの情報ニーズとしても多く示されている。例えば、2018年3月時点で教育図書館が公開しているレファレンス事例¹⁴⁾では、652件中278件が教科書に関わるもの（「教科書」で検索した事例）である。すなわち教育図書館のレファレンス質問の4割超を教科書に関わる内容が占めている。

3. 教科書 LOD の利用

教科書 LOD は単なるデータの集まり、データセットであるだけでなく、LODの提唱者 Tim Berners-Lee の提案¹⁵⁾にしたがって、教科書の URI にアクセスすれば閲覧者にとって有益な情報を提供するウェブサイトとなっている。

教科書の情報を得たい場合は、その教科書の URI にアクセスすれば、当該教科書の基本的な情報を得ることができる。例えば、図1は、URI “https://w3id.org/jp-textbook/中学校/2011/公民/921” にアクセスした画面を表示している（ただし、後述する固定 URI サービスを用いているので、実際のブラウザのアドレス欄には “https://jp-textbook.github.io/中学校/2011/公民/921” が表示されている）。URI から明らかなように、この教科書は2011年に検定された中学校の公民の教科書である。さらに、この教科書ページを見れば、書名が「新しい社会公民」、編著者が「五味文彦 戸波江二 矢ヶ崎典隆 ほか46名」、出版者が「東京書籍」、ISBN が「978-4-487-12048-2」

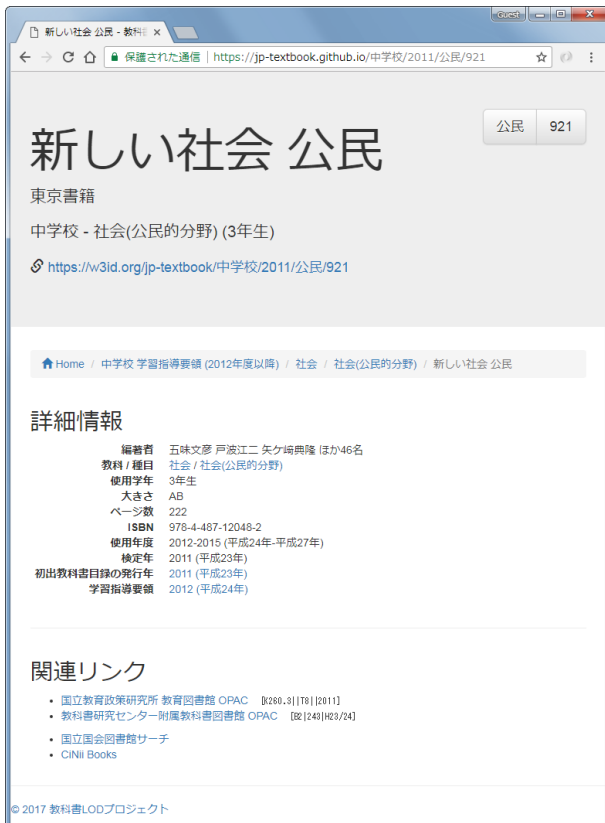


図1 教科書情報の表示例



図2 トップページを表示例

であるといった基本的な書誌情報を得ることができ、教科が「社会」、種目（科目）が「社会（公民的分野）」、使用学年が「中学校」の「3年生」、教科書が使用された年度が2012年から2015年、対応する学習指導要領が2012年版であるといった教科書特有の情報も得ることができる。また、関連リンクとして、教育図書館 OPAC（オンライン蔵書検索）の当該教科書の書誌詳細ページへのリンク、教科書図書館 OPAC の当該教科書の所蔵詳細ページへのリンクもある。さらに、国立国会図書館サーチや CiNii Books を当該教科書の ISBN で検索した結果へのリンクもある。

また、教科書 LOD のトップページ “https://w3id.org/jp-textbook/”（図2参照）には、教科書が存在する教科や種目（科目）のリンクがあり、教科や種目（科目）のリンクを使って、個々の教科書のページにたどりつけるようになっている。また、トップページの下部では、教科書 LOD 全体のデータセットを、Turtle 形式¹⁶⁾により一括ダウンロードできるようにもなっている。

また、「教科書 LOD について」として、教科書 LOD についてのデータモデルの説明や、これまでのリリースの履歴などを閲覧できるページ¹⁷⁾も用意した。このページへのリンクはすべてのページのフッターに用意しており、Google などの検索エンジンの検索結果から直接、教科書のページにたどりついた場合でも、わかりやすくだれるよう配慮している。

4. URI 設計

教科書 LOD では、教科書リソースを中心に、学習指導要領、教科、種目、教科書目録といった付随する情報それ

ぞれに URI を付与して Linked Data リソースとし、これらに関連するデータもしくはメタデータをモデル化した。

URI 設計で用いる教科書 LOD のベース URI は“https://w3id.org/jp-textbook/”を用い、以下ではプレフィックス textbook:と表記する¹⁸⁾。このアドレスは World Wide Web Consortium Permanent Identifier Working Group が運営する固定 URI サービス¹⁹⁾を用いており、実際に公開するドメイン名やアドレスが変わっても、URI が変わらないよう工夫している。なお、2018年2月現在、教科書 LOD は Github の Web ページ公開サービスを利用しており、その実体は、“https://jp-textbook.github.io/”にある。固定 URI サービスの“https://w3id.org/jp-textbook/”にアクセスすると、Github 上にある実体へ転送され、教科書 LOD の情報にアクセスできる。

以下では、教科書、学習指導要領、教科、種目、教科書目録それぞれのリソースをあらわす URI について説明する。

4.1 教科書リソースの URI

各教科書は出版された固有のタイトルを一つの単位とした。教科書は検定制度のもとで、文部科学省により使用する教科書が認定される。そのため、文部科学省が毎年発行する学校種別ごとの「教科書目録」には、次年度に学校教育で使用可能なすべての教科書の書誌情報が掲載されている。

この教科書目録では、各教科書はタイトルごとに、教科書記号と教科書番号の2つの識別子を付与して提供される。例えば、図3は、中学校の教科書目録の一部を表しているが、この表の3列目に、「教科書の記号・番号」とあり、この列に、教科書記号と教科書番号が示されている。この例では、教科書記号が「音楽」で、教科書番号が「725」や「825」などである。この教科書記号と教科書番号は、検定後の教科書の情報を整理・提供するために、文部科学省が教科書のタイトルごとに付与して利用している識別子である。第1章でも述べたように、教科書記号、教科書番号は教科書の表紙や奥付などにならずに記載されている。

この2つの識別子は当該使用年度を超えて、翌年度以降

音楽 音楽(一般)

※教科書の記号・番号欄の「※」は拡大教科書の発行予定があることを示す(はしがき参照)

発行者 の番号 ・略称	使用 学年	教科書 の記号 ・番号	書名	判型 ページ数	予 定 価 (円)	検定 済年	著者
17 教出	1	音楽 725 ※	中学音楽 1 音楽のおくりもの	AB 変型 96	245	平27	新実 徳英 ほか18名
		音楽 825 ※	中学音楽 2・3上 音楽のおくりもの	AB 変型 96	244	平27	
		音楽 826 ※	中学音楽 2・3下 音楽のおくりもの	AB 変型 96	243	平27	
27 教芸	1	音楽 727 ※	中学生の音楽 1	A4 変型 90	245	平27	小原 光一 ほか14名
		音楽 827 ※	中学生の音楽 2・3上	A4 変型 90	247	平27	
		音楽 828 ※	中学生の音楽 2・3下	A4 変型 90	240	平27	

図3 教科書目録の例(「中学校教科書目録(平成29年度使用)」[20, p.9]より引用)

も同一の識別子が付与されるため、これらに学校種別と検定年の2つを加えて、合計4つの情報を使うことによって、使用年度を超えて、教科書のタイトルごとの識別ができる。検定年、学校種別についても、奥付や表紙などに必ず教科書に記載されている。よって、教科書リソースの URI を“textbook:学校種別/検定年/教科書記号/教科書番号”として、構築することとした。

例えば、学校種別が「高等学校」、検定年が2012年、教科書記号が教科「国語総合」を示す「国総」、教科書番号が301として識別される教科書タイトルの URI は“textbook:高等学校/2012/国総/301”として設定される。

なお、原則として、教科書の内容の改訂は、検定のタイミング(概ね4年ごと)で行われるため、改訂されたものに対して同一の URI が割り当てられることは無い²¹⁾。同様に、異なる教科書には異なる教科書番号が付与されるため、異なる教科書が同じ URI となってしまうことは無い。

4.2 学習指導要領リソースの URI

日本では、初等・中等教育で教える内容については、カリキュラムなどのガイドラインとして、文部科学省によって、「学習指導要領」が開発されている。学習指導要領は、おおむね10年に一度改訂されている。教科書は、この学習指導要領の内容に沿って作成されるため、教科書 LOD では、教科書リソースに対応する学習指導要領を示す、学習指導要領の URI も決定した。

学習指導要領は、1)学校種別ごとに開発されること、2)その学習指導要領を適用し始める日付(学習指導要領施行年)が決まっていることから、この2つを使って URI とし、“textbook:curriculum/学校種別/学習指導要領施行年”と表現することにした。例えば、1994年に施行される高等学校の学習指導要領は、“textbook:curriculum/高等学校/1994”と表現される。

4.3 教科リソースの URI

教科書は、どの教科で利用されるかを想定して作成されているため、教科書に対応している教科を示すために、教科を示す URI を決定した。教科は、学習指導要領によって定義されているため、学習指導要領の URI に教科名を加えて“textbook:curriculum/学校種別/学習指導要領施行年/教科名”と表現することにした。教科名は、学習指導要領の記載に基づいて付与した。例えば、1994年に施行された高等学校の学習指導要領に記載された教科「国語」は“textbook:curriculum/高等学校/1994/国語”と表現される。

4.4 種目リソースの URI

種目は、教科書作成者が、検定時に申請する教科書の種目を指すものである。これはおおむね、学習指導要領における科目とほぼ同一のものを指すが、厳密には、科目とは異なる実体である。例えば、種目には「地図」や「書写」など、科目としては存在しないものが含まれ、「保健体育」のように、複数の科目に対応する種目もある。

どの種目の教科書かを一意に示すために、種目の URI を決定し、教科書と関連付けることとした。種目には対応する教科があるため、教科の URI に付随する形で“textbook:curriculum/学校種別/学習指導要領施行年/教科名/種目名”と表現することにした。例えば、1994 年に施行された高等学校向けの学習指導要領に記載された科目「国語表現」は“textbook:curriculum/高等学校/1994/国語/国語表現”と表現される。種目名は、文部科学省が、検定の申請者に提示している「教科用図書検定申請受理種目」という告示文書²²⁾の記載に基づいて付与した。

4.5 教科書目録リソースの URI

教科書目録とは、4.1 節で述べたように、文部科学省が毎年発行している、次年度に使用できるすべての教科書のリストである(図 3 を参照)。教科書目録は、教科書の存在を示している重要な一次資料であるため、教科書目録をリソースとして URI を付与することとした。

教科書目録は、学校種別ごとに発行され、かつ教科書目録は毎年発行されることから、学校種別と発行年を使って URI とし、“textbook:catalogue/学校種別/発行年”と表現することにした。例えば、1994 年に発行された高等学校向けの教科書目録は、“textbook:catalogue/高等学校/1994”と表現される。

5. 教科書 LOD のデータモデルの概要

図 4 は、教科書 LOD の主なリソースのモデル概要をある教科書を例として示したものである。それぞれのプロパティの詳細については後述するが、ここでは教科書 LOD 全体のデータモデルの概要を説明する。

図の楕円はリソースを、四角はリテラルを表している。真ん中の“textbook:中学校/2001/英語/904”は、教科書リソースを表しており、教科書リソースは、書名プロパティのリテラル値“NEW CROWN ENGLISH SERIES 3”に加え、出版者名などの書誌情報を表すリテラルを持つことがわかる。また、検定年プロパティのリテラル値“2001”や、使用年プロパティの“2002-2005”など、教科書に特有の情報も記述される。またこの教科書の教科を示すリソース“textbook:curriculum/中学校/2002/外国語”や、種目を示すリソース“textbook:curriculum/中学校/2002/外国

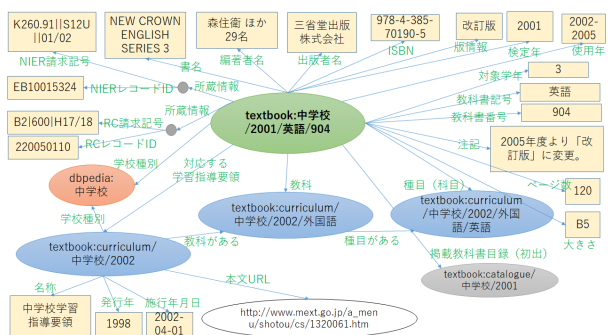


図 4 教科書 LOD の主なリソースのモデル概要

語/英語”が教科書リソースと関連付けられる。また、この教科書に対応した学習指導要領として、学習指導要領リソース“textbook:curriculum/中学校/2002”がある。教科書リソースは対応する学習指導要領リソースも示してあり、種目(科目)も対応する教科書リソースと関連付けられている。また、教科書が使用される学校種別は、DBpedia の対応する学校種別リソース“dbpedia:中学校”を関連付けている。教科書が初めて掲載された教科書目録リソース“textbook:catalogue/中学校/2001”とも関連付けている。

教科書リソースの単位としては、教科書 1 タイトルごとに、1 つの教科書リソースとなるようにしている。つまり、同じ内容の教科書はすべて、1 つの URI として指し示す。

一方で、同じ教科書を複本として持っている場合や教育図書館以外の他の図書館に所蔵された教科書を別々に指し示す必要から、個別の所蔵資料単位の情報として、所蔵資料プロパティを定義し、各所蔵資料リソースを空ノードとし、教科書リソースと関連付け、それぞれの所蔵資料の情報としては請求記号プロパティなどを通じて、個別の所蔵資料の情報を記述している。図 4 では、2 つの所蔵情報が教科書と関連付けられており、それぞれ教育図書館、教科書図書館の 2 箇所に由来する個別の所蔵資料の情報を記載している。

なお、出版社が倒産するなどして、まったく同一の内容の教科書をほかの出版社が出版するといった例外的なケースがまれに存在する。この場合、学校種別、検定年、教科書記号、教科書番号の 4 つの情報では識別できない教科書リソースになるため、注記にその旨がわかるようにした。なお、2018 年 3 月現在、7,254 件の教科書リソース中、63 リソースがこの例外的なケースに相当する。この例外的な教科書リソースについては、教育図書館は別の書誌レコードとして扱っているため、教育図書館の所蔵情報を表すために、空ノードを使って教育図書館の書誌レコード ID と請求記号をそれぞれの別リソースとして関連付けられるようにした。

6. プロパティ語彙

教科書 LOD のプロパティ語彙は、なるべく広く使われている語彙を使用するように努めた。ただし、よりよいプロパティ語彙を探す時間を掛けるよりも、まずは筆者らが持っている情報を LOD として公開することを優先する方針をとった。筆者らの見落としなどにより、より適切なプロパティ語彙を採用できていない可能性も大いにあるが、よりよいプロパティ語彙を見つけたことができれば、その都度、修正版をリリースすることで対応することとした。プロパティ語彙が独自語彙になったとしても、また、多少不自然でも、公開されていないよりは、公開されている状態がより望ましいと考えたからである。また、公開すれば、専門の有識者などに相談もしやすいという点も、このような方針の利点である。以上の方針により、教科書 LOD では 2018 年 3 月現在、次のようなプロパティ語彙を用いることとした。

書誌情報のプロパティ語彙は原則として Schema.org²³⁾ を使用した。Schema.org を採用した理由は、主要な検索エンジン企業により提唱され、ウェブ上で活発にメンテナンスされて続けているメタデータスキーマであることに加え、LOD 語彙としての対応範囲が広いこと、OCLC での採用²⁴⁾など図書館コミュニティでの利用も多いことなどが理由である。Schema.org のプロパティとしては、例えば、書名 (name) や、編著者名 (editor)、出版者 (publisher) などを扱っている。一方で、Schema.org に現時点で見当たらない書誌情報関連のプロパティとして、ページ数 (extent) や大きさ (dimensions) の項目は、BIBFRAME 語彙²⁵⁾を用いた。学習指導要領の本文や教科書目録の本文など、文部科学省サイトなどで公開される情報への参照をあらわすプロパティは、“rdfs:seeAlso”を使用した。また、教科や種目の並び順を示すプロパティには RDF Data Cube 語彙²⁶⁾から“qb:order”を用いた。

そのほか、教科書特有のプロパティについては独自語彙を名前空間 <https://w3id.org/jp-textbook/>以下に設定して、検定年 (authorizedYear)、使用年 (usageYear)、学年 (grade)、教科 (subjectArea)、種目 (科目) (subject) などの情報を記述することとした。また、教育図書館の所蔵情報を表すプロパティも独自語彙として名前空間 “<http://dl.nier.go.jp/library/vocab/>” を、教科書図書館の所蔵情報として名前空間 “<http://dl.nier.go.jp/library/vocab/textbook-rc/>” を使用した。各リソースのプロパティの詳細については、教科書 LOD のサイト上の文書¹⁷⁾を参照してほしい。

7. LOD データへの変換

LOD データモデルはグラフモデルであり、既存のデータベースでの管理形式とはデータモデルが異なるため、一括で変換する必要がある。本章では、この変換処理について説明する。基本的には、表形式で管理されているデータをグラフモデルの LOD の標準的な形式 (具体的にはもっとも普及している Turtle 形式) に変換した。教科書の書誌・所蔵データから LOD に変換するまでのデータフローを図 5 に示す。

教科書リソースの主な書誌情報については、教育図書館で作成された書誌データが元となっている。まず、教育図

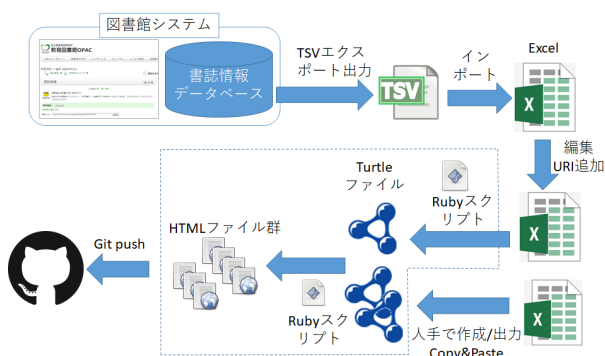


図 5 教科書 LOD 構築のデータフロー

書館で作成されたデータを TSV 形式 (Tab Separated Values 形式) で提供してもらった。次に、この TSV 形式のファイルを、Excel にインポートした。対応する学習指導要領については、教育図書館のデータには付与されていないため、筆者らが初出の教科書目録や検定年などから調査し、人手で学習指導要領リソースの URI を付与した。最後に、筆者らが Ruby で開発したプログラムによって、この Excel ファイルから Turtle 形式に変換した。

教科書の ISBN は、教育図書館の書誌データには付与されていないため、教科書図書館で作成されたデータを入力して、Excel ファイル上で関連付けて変換処理に加えた。

作成した Turtle 形式のファイルをウェブ上で公開すれば LOD データセットの公開となるが、LOD においては、リソースに対応するウェブページの公開が必要となる。これは、LOD 原則の 3 点目「リソースの URI を参照したら、閲覧者に有益な情報を提供すること」¹⁵⁾ という方針に沿って、各リソース URI に対応する情報を各 URI にアクセスしたときに提供するものである。教科書 LOD では、Turtle 形式のファイルに対して、独自に開発した Ruby プログラムを用いて、各リソースの情報を表現する静的な HTML ファイルに変換した。変換した HTML ファイル群は Github Pages²⁷⁾の機能を用いて、Github 社が提供するクラウド上のウェブ公開領域で公開した。さらに、この Github Pages の実体アドレスに対して、先述した固定 URI サービスからリダイレクト処理を行うことにより、Github 以外のサービスに公開場所を移動しても問題ないような形とした。

8. これから LOD を作成、公開する人のために

最後に、教科書 LOD の開発と公開を通じた学びとして、他のデータセットの開発や公開にどのような点が役立つかという視点から、いくつかの点に触れておきたい。

8.1 URI の設計で考慮すべきこと

URI は識別子 (ID) であるので、一意に識別できるかをまず初めによく検討する必要がある。ただし、この一意性も、対象物がどのようなニーズに照らして求められる情報か、どの程度までの正確性が求められるかを検討して定めるとよい。たとえば、教科書 LOD では、教科書の 1 つのタイトルが別のリソースにならないよう、また、異なる教科書が同一の URI を共有することのないよう注意した。この際、全てのレコードをまとめて検討すると、倒産した出版社や著作権譲渡したケースなどでいくつかの例外があることに気づいた。このような例外は少なくとも 30 年分の教科書コレクションの中では相当に例外的なケース (7,254 件中 63 件=0.9%弱) として無視しても差し支えない分量と判断して、特にこれらの教科書同士の違いについて URI 上は区別せず、注記プロパティを設けて説明するだけにとどめることとした。

LOD の世界では URI は永続的な識別子であるので、外部的要因によって変化することは好ましくない。たとえ

ば、国の組織は数十年単位で見れば、組織改編等により、組織が丸ごと別の形態に変化することも容易に想像できる。このため、教科書 LOD では、世界的によく使われている固定 URI サービスを使って、実体としてのアドレスは変わっても、LODの世界でリソースとして表現される URI は変化しないようにした。なお、教科書 LOD では固定 URI サービスとして w3id.org を用いたが、他にインターネットアーカイブが運営する purl.org サービスもある²⁸⁾。

さらに、各 URI を設計する基本方針として、URI を見ただけで、そのリソースがどのようなものかが想像できるようなものが望ましい。たとえば教科書 LOD では、教科書の表紙等に書かれている教科書記号や教科書番号、学校種別をもとに URI を表現することによって、URI を見ただけでも科目や学校の種類がある程度までは分かるように配慮した。

8.2 サービスとしての継続性

LOD 世界では URI リソースの永続性が問われるのと同様、データ提供サービスとしての継続性も重要である。LOD データの提供も他のサービスと同様に、新しいレコードが加わるときにどのような処理が必要か、その手間はどれくらいか、そのような処理を誰が実行するかを見積もっておかなければならない。教科書 LOD は研究開発プロジェクトではあるが、継続性を意識して、教育図書館や教科書図書館の司書との連携による、元データの更新と同期処理を意識してプロジェクトを進めた。責任の範囲として、データ内容そのものとデータの一括出力処理は、教育図書館や教科書図書館の司書側が対応し、筆者らは、それらのデータ群を渡されたあとの LOD 変換の処理に注力するという責任分担を設けた。これにより、データ内容に誤りがあった場合にも、その連絡、応答を通じて元となる書誌データにフィードバックすることができた。

また、教科書 LOD を作成する過程で分かってきたことは、LOD データセットを作成することの副産物的な意義はデータクリーニングが自然と発生することである。当然のことながら、既存の書誌レコード内の記述内容に誤りがあると、とりわけ URI パラメータに用いるようなデータ項目では、Linked Data の関連付けリンクが機能しなくなり、リソース同士の対応付けがとれないことになる。こういった処理は、教科書 LOD のデータセットを処理するなかでいくつも発見し、そのたびに、教育図書館や教科書図書館に依頼して直してもらうことにした。これにより、教育図書館や教科書図書館のルーチン的な作業で構築してきたデータセットがクリーニングされ、より正確なデータとして利用者に提供できることは別の角度の利点であると実感している。

9. おわりに

本稿では教科書 LOD の開発および公開の際に検討した論点を振り返りながら、得られた知見を紹介した。教科書 LOD は、7 千件程度とさほど大きくない規模ながら、図書

館が扱う資料の書誌情報と関連情報をまとめて構造化し、LOD として扱いやすいように URI リソースを設計して共有できるようにしたものである。教科書 LOD は教科書のメタデータおよび教科などの情報を CC0 として自由に使えるライセンスにより公開したものであり、さまざまな形でぜひ活用いただきたい。従来、図書館を含む多くの情報機関では、構造化された情報を利用者が求める形で提供できるよう組織化してきた。今後は、LOD を含む多様な環境に対しても、これまでの資産を活かして展開していくことがますます求められると考える。本稿がそのような多様な展開の一助になれば幸いである。

教科書 LOD の今後の課題には、目次や全文レベルの情報を含む教科書内容への展開、さらなる各地の図書館所蔵資料への対応、所蔵資料への同定識別子の導入、NCID や JPNO などの国内の書誌識別子との対応付けなどがある。今後、地道な改善と情報提供を継続するとともに、こういった有用性を高めるための機能追加も検討していきたい。

謝辞

国立教育政策研究所教育図書館及び教科書研究センター附属教科書図書館には、教科書 LOD の元となる教科書の書誌データの提供をいただきました。ここに感謝の意を表します。

註・参考文献

- 1) 特集：つながるデータ。情報の科学と技術, vol.67, no.12, pp.613-638, 2017.
- 2) 特集：リンクト・オープン・データの利活用。情報処理, vol.57, no.7, pp.586-625, 2016.
- 3) 高久雅生。知識をリンクする技術。図書館情報学を学ぶ人のために、逸村裕、田窪直規、原田隆史（編）、第 5 部第 19 章, pp.201-213。世界思想社, 2017.
- 4) トム・ヒース、クリスチャン・バイツァー。Linked data: Web をグローバルなデータ空間にする仕組み。近代科学社, 2013, pp.1-139.
- 5) 橋詰秋子、福山樹里、リンクト・オープン・データの利活用：4。出版物に関するメタデータと国際書誌コントローラー-国立国会図書館における LOD の取り組み。情報処理, vol.57, no.7, pp.606-611, 2016.
- 6) Thomas Baker. et al. Library Linked Data Incubator Group Final Report, 2011. <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/> (参照 2016 年 8 月 25 日)。
- 7) Thomas Baker. et al. (国立国会図書館電子情報部電子情報流通課標準化推進係、田辺浩介訳)。図書館 Linked Data インキュベータグループ最終報告書, 2012., <http://www.ndl.go.jp/jp/aboutus/standards/translation/XGR-ld-20111025.html> (参照 2016 年 8 月 25 日)。
- 8) Linked Open Data チャレンジ 2016 実行委員会。Linked Open Data チャレンジ Japan 2016 受賞作品発表。 <http://lodc.jp/2015/concrete5/blog/2017-02-17> (参照 2018 年 2 月 17 日)。
- 9) 正確には「IRI (Internationalized Resource Identifier)」と呼ばれる国際化された URI 記法。教科書 LOD では URI 内に日本語文字を含んでいるため、国際化されていることは特に重要である。
- 10) 文部科学省。学習指導要領「生きる力」：現行学習指導要領（本文、解説、資料等）。 http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/index.htm (参照 2017 年 10 月 10 日)。

- 11) 文部科学省初等中等教育局. 教科書制度の概要.
http://www.mext.go.jp/a_menu/shotou/kyoukasho/gaiyou/04060901.htm (参照 2017 年 10 月 10 日).
- 12) 文部科学省初等中等教育局. “付表 3 教科書の種類数・点数・需要数”. 教科書制度の概要, 2017.
http://www.mext.go.jp/a_menu/shotou/kyoukasho/gaiyou/04060901/1235103.htm (参照 2017 年 10 月 12 日).
- 13) 国立教育政策研究所国際研究・協力部. OECD 生徒の学習到達度調査 (PISA). <http://www.nier.go.jp/kokusai/pisa/> (参照 2017 年 10 月 10 日).
- 14) 国立教育政策研究所教育図書館. 教育図書館のレファレンス事例検索 (レファレンス事例集).
http://www.nier.go.jp/library/ref_search.html (参照 2018 年 2 月 17 日).
- 15) Tim Berners-Lee. Linked data.
<https://www.w3.org/DesignIssues/LinkedData.html> (参照 2017 年 11 月 29 日).
- 16) David Beckett. et al. RDF 1.1 Turtle: Terse RDF Triple Language, 2014. <https://www.w3.org/TR/turtle/> (参照 2018 年 3 月 14 日).
- 17) 教科書 LOD について. <https://w3id.org/jp-textbook/about> (参照 2017 年 10 月 10 日).
- 18) プレフィックスとは接頭辞のことであり, URI を省略して記載したい場合に用いる。例えば, ある URI “<https://w3id.org/jp-textbook/学校種別/検定年/教科書記号/教科書番号>” における冒頭のベース URI 部分を共通化し, 短い文字列, つまり接頭辞のみで表現できる。この URI をプレフィックス付きで表現すると URI “textbook:学校種別/検定年/教科書記号/教科書番号” と示せる。本稿の場合, “textbook:” は <https://w3id.org/jp-textbook/>” と読み換えることとする。
- 19) W3C Permanent Identifier Community Group. Permanent identifiers for the web. <https://w3id.org/> (参照 2017 年 10 月 10 日).
- 20) 文部科学省. 中学校用教科書目録 (平成 29 年度使用).
http://www.mext.go.jp/a_menu/shotou/kyoukasho/_icsFiles/afieldfile/2016/04/20/1369944_2.pdf (参照 2017 年 10 月 10 日).
- 21) ここで「原則として」としたのは, 例えば, 新元素「ニホニウム」の発見のような新事実の発見などにより教科書の内容を変える必要ができたときなど, 次の検定を受ける機会を待たずに, 内容に関わる部分で修正される教科書があるためである。その場合は文部科学省に訂正申請を行って承認されると検定を経ずに修正できる。
- 22) 文部科学省. 教科書検定の申請受付.
http://www.mext.go.jp/a_menu/shotou/kyoukasho/kentei/shinsei.htm (参照 2018 年 4 月 9 日).
- 23) schema.org. <http://schema.org> (参照 2018 年 2 月 22 日).
- 24) OCLC WorldCat.org が Schema.org によるマークアップを採用, Linked Data へ向けた取組の一環として.
<http://current.ndl.go.jp/node/21161> (参照 2018 年 2 月 22 日).
- 25) BIBFRAME 2.0 Vocabulary.
<http://id.loc.gov/ontologies/bibframe/> (参照 2018 年 3 月 1 日).
- 26) The RDF data cube vocabulary (W3C recommendation).
<https://www.w3.org/TR/vocab-data-cube/> (参照 2018 年 3 月 1 日).
- 27) What is GitHub Pages?
<https://help.github.com/articles/what-is-github-pages/> (参照 2018 年 3 月 1 日).
- 28) PURL Administration. <https://purl.org> (参照 2018 年 5 月 24 日).

Case Study: Building and Publishing Japanese Textbook Linked Open Data. Yuka EGUSA (Office for Educational Resources Research Promotion, Department of Research Planning and Development; National Institute for Educational Policy Research, 〒100-8951 3-2-2 Kasumigaseki, Chiyoda-ku, Tokyo, 100-8951, Japan), Masao TAKAKU (Faculty of Library, Information and Media Science; University of Tsukuba)

Abstract: Japanese Textbook LOD is a dataset, which is compiled from a collection of textbooks held and organized by NIER Education Library and Textbook Research Center Library for several decades. The dataset includes the textbooks published up to now since the 7th curriculum guideline enforced in 1992, as well as the related information such as the subject areas, the subjects, and the curriculum guidelines. As of March 2018, the dataset contains bibliographic information for 7,257 textbook titles and 157,297 triples as RDF data. This paper describes findings and lessons learned from the development and publishing the dataset.

Keywords: Textbook / Linked Open Data / bibliographic information / Uniform Resource Identifier (URI)