

S×UKILAM 教材アーカイブの LOD 化： RDF と SPARQL によるデジタルアーカイブを活用した教材と 多様な教育情報の接続・構造化

大井 将生（人間文化研究機構／東京大学）

中村 覚・大向 一輝・渡邊 英徳（東京大学）

概要：本研究の目的は、DA 資料を用いた教材と多様な教育情報を機械可読性の高い形で接続・構造化する LOD モデルを構築することである。そのために、S×UKILAM 教材アーカイブのデータを用いた RDF データセットと SPARQL エンドポイントを開発する。その結果、教材と学習指導要領コードやジャパンサーチ・NHK の動画コンテンツなどの教育情報を機械可読性の高い形で接続し、検索可能な LOD モデルを構築することができた。これにより、DA 資料を用いた教材と多様な教育情報を紐付けた検索が容易となり、教育の情報化とデジタル文化資源の活用促進に貢献することができたと考える。

キーワード：LOD, RDF, SPARQL, S×UKILAM 連携, 学習指導要領コード, デジタルアーカイブ

Converting S×UKILAM learning materials archive for LOD: Connect and structure educational materials and diverse educational information using digital archives by RDF and SPARQL

Masao OI (National Institutes for the Humanities / The University of Tokyo)

Satoru Nakamura / Ikki Ohmukai / Hidenori Watanabe (The University of Tokyo)

Abstract : The purpose of this study is to construct an LOD model that connects and structures educational materials by utilizing DA resources and a variety of educational information in a machine-readable form. For this purpose, we develop an RDF dataset and a SPARQL endpoint using data from the S×UKILAM archive. As a result, we are able to construct a findable LOD model that connects educational materials with diverse information such as Courses of Study codes, Japan Search and NHK video contents in a highly machine-readable form. We consider that this model facilitates the search of educational materials using DA resources and various educational information and contributes to the informatization of educational and the enhancement of the use of digital cultural heritage.

Keywords: LOD, RDF, SPARQL, S×UKILAM collaboration, Japanese Courses of Study Code, Digital Archive

1. はじめに

社会の情報化が進む中で、デジタルアーカイブ（以下 DA）の構築と活用の需要が高まっている。とりわけ、GIGA スクール構想下で多様な資料を用いた学習の重要性が増した学校教育に資する情報基盤の整備は重要な課題である。さらに、今後デジタル教科書等のデジタル教育環境の急速な普及が予想される中においては、Web 上に散在する様々な資料・情報とカリキュラムなどの教育情報とを有機的に接続する需要が高まっている。

「デジタルアーカイブ憲章」（2023）においても、多様な関心に応える学びの基盤を構築し、学習の質を高めると共に情報リテラシーの向上や歴史的・国際的な視点を育むことが目指されている[1]。こうした基盤を整備し、デジタル文化資源の情報を授業で活用可能な環境を構築することで、学習者の意欲や学習効果を向上させることができることも明らかになっている[2]。

一方で、DA 資料と学校教育における学びとを関連づける体系的な方法は確立されていない[3]。また、DA 資料の教育活用の方法を共有する仕組みも十分に検討されていない。教育現場の目線に基づいた DA 資料の活用を促進するためには、以下の要件を満たす手法を検討する必要がある。

- DA 資料と教材（活用方法）やカリキュラムなどの教育情報との接続
- DA 資料への教育現場の目線に基づいたメタデータ（以下「教育メタデータ」）の付与
- DA 資料と教材の関係性の把握や関連情報の検索を容易にするための構造化されたメタデータの整備

そこで本研究は、「DA 資料を用いた教材と多様な教育情報を接続・構造化する Linked Open Data（以下 LOD）モデルを構築すること」を目的とする。

2. 関連研究

2.1 DA 資料への教育メタデータの付与と課題

Web 上のコンテンツに対する教育現場の目線に基づいた検索が可能である代表的なサービスとしては「NHK for school」[4]があり、教科書や学習指導要領から動画コンテンツを検索することができるため、広く教育現場で活用されている。

DA 資料とカリキュラム等の教育情報を接続する方法としても、資料に校種・学年・単元・教科などの教育現場の目線に基づいたメタデータ（以下「教育メタデータ」）を付与することが考えられる。「教育メタデータ」に基づいた検索が可能な DA は稀有であるものの、福井県文書館の「学校向けアーカイブズガイド」[5]が先進的な事例としてあり、2023 年 11 月時点で 62 件の資料が単元・校種ごとに検索可能である。

しかしながら、「教育メタデータ」の付与には教育学的な知識や経験が求められるため、資料公開機関が単独では実行しづらいという課題がある。さらに、学校教員や児童生徒が DA 資料を実際に活用する際には、「問い」に即した自由な発想で、単元などの固定的な要素に縛られずに検索を行いたいという需要がある。そのため、キーワードなどを含めた多様な観点での「教育メタデータ」付与とその活用方法のあり方について検討を進める必要がある。

2.2 共創的な教材化と教育メタデータの付与

資料公開機関と学校教育を架橋する試みとしては、「S×UKILAM（スキラム）連携」が挙げられる[6]。このスキーマでは、小中高の教員や資料公開機関の関係者らが共創的に DA 資料を教材化し、「教育メタデータ」を DA 資料自体に付与するのではなく、「加工された」教材に対して付与している点に特徴がある。S×UKILAM 連携で共創される教材は、教科・単元・時代・キーワード・視点・授業場面・育みたいコンピテンシー・評価指標・学習指導要領コードなど、多様な観点で付与された「教育メタデータ」と共に、包括的に閲覧可能な形で IIIF ビューアを用いて公開されている[7]。

しかしながら、上記で示した公開データの状態では、教材と他のコンテンツや関連情報との接続は為されていないという課題が残る。限られた時間の中で教材開発や探究学習を行うためには、カリキュラムに即して構造化された様々な関連資料や情報にシームレスにアクセス可能な状態であることが望まれる。

2.3 DA に関する LOD

上述した課題を解決するためには、DA 資料を用いた教材と多様なコンテンツや関連情報を機械可読性の高い形で接続・構造化するための LOD データモデルを構築することが有効であると考えられる。林ら（2016）は、Web 上の情報資源の多く

が PDF など機械可読な構造化が不十分であることをふまえて、異なる性質の要素を持つ複合的な情報資源に対して相互運用性を持つ LOD の適用を検討することの重要性を指摘した[8]。また嘉村ら（2017）は、情報の構造が異なる資料群を総合的に情報システムとして扱うためには、事後的に情報を追加することができる半構造型で表現の自由度が高いグラフモデルを用いることが望ましいとしている[9]。

そこで本研究では、主語・述語・目的語を連ねることで有向グラフを構成し、機械可読性の高い構造化データを表現が可能である Resource Description Framework（以下 RDF）を用いた LOD モデルの検討を行う。

DA に関連する LOD の代表的な取り組みとしては、Dublin Core 2011 で発表されたパイロット[10]を元に Europeana が 2012 年から行っている大規模データの LOD 化[11][12]が先駆的なモデルとして挙げられる。国内でも、武田らによって 2012 年より DBpedia Japanese が公開され、分野連携型データベースの枠組みが検討されてきた[13]。2020 年には国の分野横断統合プラットフォームであるジャパンサーチにおける利活用のためのメタデータスキーマ LOD として「ジャパンサーチ利活用スキーマ」[14][15]が公開されている。その設計に際しては利用者タスクの観点で扱いやすく連携元にもメリットがあることが目指されている点は参照性が高い。ジャパンサーチ LOD では外部の RDF の活用が目指されており[16]、RDF を媒介としたジャパンサーチと接続可能な教育コンテンツの拡充が望まれる。

2.4 教育情報に関する LOD

教育分野に関連する LOD 化の取り組みとしては、「教科書 LOD」[17]や「学習指導要領 LOD」[18]などが RDF を用いたデータセットとして公開されている。「教科書 LOD」は初等中等教育の現行使用教科書の出版情報・単元情報・編修趣意書情報などを LOD 化した点[19]、「学習指導要領 LOD」は文部科学省が公開している学習指導要領と教育要領の内容・コードおよび関連する情報を LOD 化した点[20]に特徴がある。両者は日本の教育現場で重視される教科書や学習指導要領をデータモデルの中心に位置づけてデータを設計している点で参照性・応用性が高く、教育分野における基盤情報の LOD モデルとして有用である。

本研究においても、これらの既存の取り組みを参考にするとともに、既存のデータセットとリンク可能な LOD モデルを検討する。中村ら（2018）は、DA の活用に際しては、ユーザの目的に応じた史料の収集・整理・分析を可能とする環境が重要であることを指摘している[21]。このことをふまえて、学校における DA の活用を対象とした本研究では、教育現場の需要や目的に即した視点でデータモデルを検討する。

さらに、以下のクエリより、「教育メタデータ」のキーワードに紐づくジャパンサーチ上のキーワード（及び人物）が 622 件接続されていることが確認できる。

```
select distinct ?jpsKeyword WHERE {  
  ?s semp:キーワード ?keyword. ?keyword  
  rdfs:seeAlso ?jpsKeyword . }
```

このようなジャパンサーチとの接続により、「富士山」「防災」「徳川家康」「第二次世界大戦」など、教材開発時に参照性の高い一般的な語彙をもとにジャパンサーチの様々な資料から教材を検索・参照可能になったこと点は有用性が高いと考える。ジャパンサーチは 2023 年 11 月時点で日本各地の 222 連携データベースから 2900 万件を越える多様な資料のメタデータを横断検索可能であり[28]、今後の教育現場における情報基盤としての役割が期待される。本研究で示した LOD モデルは、今後更に拡張していくことが予想される国の分野横断型 DA プラットフォームの活用促進に資するものであると考える。

5.1.2 「学習指導要領 LOD」との接続

学習指導要領コードを媒介とした教材や関連情報の検索が可能となったことも、開発した LOD の利点として挙げられる。「学習指導要領 LOD」との関連付けを行ったことで、SPARQL エンドポイント上で「学習指導要領 LOD」の URI を介して教材に関する指導要領の説明文の情報を取得できるようになった。これにより、指導要領の内容から教材を検索することや、同じ指導要領項目でも異なる視点や学習デザインで制作された教材を発見することも可能になった。同様に、学習指導要領コードが付与されている NHK の動画コンテンツ情報などの既存の教育系の情報／資料を接続できるようにしたことは学校現場の需要に応えるという点で有用性が高いと考える。

例えば、ワークショップで付与された学習指導要領コードは、教材アーカイブでの公開段階では「8322203121100000」のような数字の羅列に過ぎず、それが何を示しているのか把握できないという課題があった。これに対して、開発した LOD を活用することで「中学校」の「社会科」の授業において、以下に示す学習指導要領の教育目標に即して制作された教材であることを自然言語で把握できるようになった。

「自らが生活する地域や受け継がれてきた伝統や文化への関心をもって、具体的な事柄との関わりの中で、地域の歴史について調べたり、収集した情報を年表などにまとめたりするなどの技能を身に付けること」

5.1.3 開発した LOD の利点に関するまとめ

以上の結果は対象データセットを LOD 化した利点を示すものであり、ジャパンサーチなどの DA 資料や学習指導要領・NHK の動画コンテンツなどの教育情報を学習教材と紐付けて検索することを容易としたことで、Web 上に散在する多様な資料や情報と学校教育を架橋する LOD モデルの一例を示すことができたと考えられる。

本研究の成果は、教育の情報化とデジタル文化資源の活用を促進し、今後急速に普及が進むことが予測されるデジタル教科書やデジタル副読本において各地の文化資源をカリキュラムに接続させて活用することに貢献し得る。

5.2 課題と展望

5.2.1 ユーザビリティ

本研究の課題として、SPARQL エンドポイントを操作するためには、ある程度 SPARQL 言語に習熟する必要があるため、本研究で構築した Snorql を直接的に活用することが難しいユーザが一定数いることが考えられる。今後は、DA 資料と教育情報を接続・構造化してネットワークで繋げる LOD の利点をエンドユーザに有用性の高い形で還元するためのアプリケーションの開発を検討する。

現在は、教材に紐づけられた学習指導要領コードを媒介として NHK for school API を活用した関連コンテンツの取得や、CiNii Books、J-STAGE 等との関連付けによる書籍・論文などの多様な情報と DA 資料を紐付けて教材の検索が可能とするアプリケーションの開発を進めている。また、本研究の手法を応用し、デジタル教科書の単元との接続や、学習指導要領の解説などとの機械可読なデータ接続についても検討する。

5.2.2 国際連携

本研究の対象データは現状では国内の資料が大半であるものの、「歴史総合」の必修化などを背景として、今後は世界中の資料やコンテンツが教材化の対象となることが予想される。その際、教材作成におけるノウハウやメタデータスキーマについて、国際的に共通化すべき点/各国でローカライズすべき点に関する議論や知識共有の体制を進展することも望まれる。そのため、今後は Europeana などの国外の DA ポータルとの連携への需要も高まることが予想される。

S×UKILAM 連携では Europeana 本部で WS を行うなど、DA 資料の教材化が国や文化を越えて同じスキーマで実現できることを明らかにしている。本研究の対象データにおいても、日本の教材で Europeana の資料が活用されている事例や欧州の Europeana メンバーが活用した教材にジャパンサーチの日本資料が活用された事例についても LOD データとして接続・構造化することがで

きた。一方で、対象データ数はまだ少なく、LODの利点や国際的な発展性に寄与するデータ構造の精緻化は十分に検討できていない。今後は国際的なワークショップ実践の蓄積と並行してデータを再構成し、LODの利点を世界的な視座で発揮できるデータモデルを構築する。

5.2.3 正規化

本研究の RDF データでは、現状では一部正規化が不十分であることも課題として挙げられる。例えば、「伊能図」と「大日本沿海輿地全図」、「月」と「Moon」などの表記揺れリソースを確認している。今後はこれらの正規化アルゴリズムを改善するとともに、オーサリングツールなどを利用してデータ作成を行う際に正規化された値を入力できるような環境構築を行う。

6. おわりに

本研究の目的は、DA 資料を用いた教材と多様な教育情報を機械可読性の高い形で接続・構造化する LOD モデルを構築することであった。そのために、S×UKILAM 教材アーカイブのデータを用いた RDF データセットと SPARQL エンドポイントを開発した。その結果、教材と学習指導要領コードやジャパンサーチ・NHK の動画コンテンツなどの教育情報を機械可読性の高い形で接続し、検索可能な LOD モデルを構築することができた。これにより、DA 資料を用いた教材と多様な教育情報を紐付けた検索が容易となり、教育の情報化とデジタル文化資源の活用促進に貢献することができたと考える。

高橋 (2021) も指摘しているように、今後はより多くの資料のデジタル化とオープン化が進むことで新しい利用を生み、新たな知が創出される基盤が形成されることが望まれる[29]。そのためには、より多くの Open Data が必要であり、資料保有機関の協力が不可欠となる。とりわけ、多様な資料の教育活用を支援する LOD を拡充するためには、学校と資料保有機関との連携を進展させることが肝要となる。このことをふまえ、多様な DA 資料を子どもたちの豊かな学びに繋げるための「人」と「データ」のネットワーク構築を実践とデータ構築の両面から進めたい。

謝辞

本研究は 2022 年度国立情報学研究所公募型共同研究(22FC03)の助成を受けて実践した。

参考文献

[1] “デジタルアーカイブ憲章”。
<https://digitalarchivejapan.org/advocacy/charter/>, (参照 2023-11-02)。

[2] V. Ferrara, A. Macchia, S. Sapia and F. Lella, "Cultural heritage open data to develop an educational framework," IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, Chania, Greece, 2014, p.166-170.

[3] 小森一輝. 学校教育におけるデジタルアーカイブ利活用のために. デジタルアーカイブ学会誌, 2019, Vol. 3 No. 2, p.211-212.

[4] “NHK for school”.

<https://www.nhk.or.jp/school/>, (参照 2023-11-02).

[5] “学校向けアーカイブズガイド”.

<https://www.library-archives.pref.fukui.lg.jp/bunsho/category/gakkoushiryou/912.html>, (参照 2023-11-02).

[6] Masao Oi, Boyoung Kim, Hidenori Watanabe. "S × UKILAM" collaboration to connect local digital resources and school education: Workshop and Archiving to construct network of "people" and "data". ICADL 2022, Vol. 13636, p.125-134.

[7] “S×UKILAM 連携:多様な資料を活用した教材アーカイブ”.

<https://adeac.jp/adeac-lab/top/SxUKILAM/index.html>, (参照 2023-11-02).

[8] 林賢紀, 瀬尾崇一郎, 阪口哲男. 複合的な情報資源の Linked Open Data 化における構造分析とスキーマ定義手法. 情報知識学会誌, 2016, Vol. 26, No. 1, p.11-28.

[9] 嘉村哲郎, 大向一輝. 人文科学における Linked Open Data の活用. 人工知能, 2017, Vol. 32, No. 3, p.401-407.

[10] Haslhofer B, Isaac A. data.europeana.eu: The Europeana Linked Open Data Pilot. International Conference on Dublin Core and Metadata Applications, 2011, p.94-104.

[11] “Linked Open Data | Europeana”.

<https://pro.europeana.eu/page/linked-open-data>, (参照 2023-11-02).

[12] Isaac Antoine, Haslhofer Bernhard. Europeana Linked Open Data – Data.europeana.eu, Semantic Web, 2013, Vol. 4, No. 3, p.291-297.

[13] 武田英明, 加藤文彦, 大向一輝. Linked Data による分野連携型データベースの枠組み. 情報知識学会誌, 2015, Vol. 25, No. 4, p.283-290.

[14] “ジャパンサーチ利活用スキーマ”.

<https://www.ndl.go.jp/jp/dlib/standards/jpsformat.html>, (参照 2023-11-02).

[15] 神崎正英. ジャパンサーチ利活用スキーマの設計と応用. デジタルアーカイブ学会誌, 2020, Vol. 4, No. 4, p.342-347.

[16] 町屋大地, 伊藤実, 高橋美知子. ジャパンサーチの LOD : デジタルアーカイブが「つながる」ために. デジタルアーカイブ学会誌, 2022, Vol. 6, No. S2, p.s82-s85.

[17] “教科書 LOD”.

<https://jp-textbook.github.io/>, (参照 2023-11-02).

[18] “学習指導要領 LOD”.

<https://jp-cos.github.io/>, (参照 2023-11-02).

[19] 江草由佳, 高久雅生. 教科書 Linked Open Data (LOD) の構築と公開. 情報の科学と技術, 2018, Vol. 68, No. 7, p.361-367.

- [20] 榎本聡, 大井将生, 高久雅生, 阿児雄之, 有山裕美子, 江草由佳. 学習指導要領の Linked Open Data 化による学習への利活用に向けた検討. 日本教育工学会研究報告集, 2022, No. 1, p.135-142.
- [21] 中村覚, 大和裕幸, 稗方和夫, 満行泰河. Linked Data とデジタルアーカイブを用いた史料分析支援システムの開発. デジタル・ヒューマニティーズ, 2018, Vol. 1, p.29-43.
- [22] “w3id.org - Permanent Identifiers for the Web”. <https://w3id.org>, (参照 2023-11-02).
- [23] “sukilam-educational-metadata”. <https://github.com/sukilam-educational-metadata/sukilam-educational-metadata.github.io>, (参照 2023-11-02).
- [24] 神崎正英. Linked Data とデータマッピング. 人工知能学会誌, 2012, Vol. 27, No. 2, p.163-170.
- [25] “Dydra”. <https://dydra.com/>, (参照 2023-11-02).
- [26] “Snorql for Japan Search について”. https://www.kanzaki.com/works/ld/jpsearch/snorql_1db-about, (参照 2023-11-02).
- [27] “Snorql for 教育メタデータ”. <https://sukilam-educational-metadata.github.io/snorql/>, (参照 2023-11-02).
- [28] “JAPAN SEARCH 現在のデータ”. <https://jpsearch.go.jp/stats>, (参照 2023-11-02).
- [29] 高橋菜奈子. 小倉百人一首 LOD のデータモデル設計と構築. 情報の科学と技術, 2021, Vol. 71, No. 8, p.376-381.