

令和 2 年 6 月 29 日現在

機関番号：82619

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H01832

研究課題名(和文)文化財情報資源の探索と発見のためのデータ連携に関する研究

研究課題名(英文) Research on data collaboration for search and discovery of cultural property information resources

研究代表者

村田 良二 (Murata, Ryoji)

独立行政法人国立文化財機構東京国立博物館・学芸企画部・室長

研究者番号：50415618

交付決定額(研究期間全体)：(直接経費) 11,500,000円

研究成果の概要(和文)：国立博物館所蔵品統合検索システム「ColBase」上の文化財のデータについて、Linked Data化するためのメタデータ語彙を検討し、CIDOC CRMの語彙を基本として独自に語彙を追加し、マッピングすることができた。このマッピングに基づいて実際にデータをRDFで出力し、SPARQL Endpointに格納してSPARQLによって検索できるようにした。作品データに含まれる作者名を索引化し、既存の典拠データと関連付けを行った。このようにして用意したデータを、既存の外部SPARQL Endpointのデータと組み合わせで連鎖的に情報を検索するマッシュアップエンジンを開発した。

研究成果の学術的意義や社会的意義

博物館や美術館の所蔵する文化財に関する情報について、他の分野の様々な情報と結びつけ、それらを連鎖的にたどりながら情報の探索や発見が可能となる環境に向けた研究を行った。こうした情報探索を行うためにどのようなデータの表現方法が最適であるか、既存の標準を検討したうえで修正を加えて、フォーマットを確定した。また実際にデータを変換して、複雑な問い合わせ条件で検索できるようにした。作品データに含まれる作者名の索引を作成し、既存の人名データとリンクさせた。これらのデータを、既存の他分野のデータとの間で連鎖的に検索できるシステムを開発した。

研究成果の概要(英文)：This research investigated metadata vocabularies for cultural properties information on "ColBase" (Integrated Collections Database of the National Museums, Japan) to transform them into Linked Data, and decided to use CIDOC CRM vocabularies as the basis and created some custom vocabularies. With this vocabularies, this research transformed data on "ColBase" into RDF and import the data into an RDF Storage and make them accessible through SPARQL queries. The research also create artists names index from "ColBase" with links to existing authority files. The research developed a mash-up engine which executes chain searches through linked data stored in external SPARQL Endpoints.

研究分野：博物館情報学

キーワード：デジタルアーカイブズ Linked Data RDF セマンティック・ウェブ 文化財情報

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年、多くの美術館や博物館が所蔵品データベースを公開しており、また複数の所蔵機関のデータを集約した統合データベースもいくつか登場している。しかし、それぞれのデータは相互に連結されておらず、検索結果により得られた情報から次の情報探索へとつなげるためには、改めて検索しなおす必要があった。Web上のデータを相互に結びつける技術を用いて、これらの情報を有機的に関連付けることができれば、各館が公開している文化財情報の価値を高めることができると考えられた。

文化財情報の記述法やデータ項目については、欧米を中心とする各国においても様々な標準が開発され、また ICOM (国際博物館会議) によるものをはじめ、国際標準の開発も進められている。我が国では、2005年に東京国立博物館において「ミュージアム資料情報構造化モデル」が研究代表者及び分担者を含む調査研究チームによって開発、公開されている(村田良二「ミュージアム資料情報構造化モデル」の構築、MUSEUM、No.602、東京国立博物館、2006年)。また同年以来、東京国立博物館では先進的な所蔵品管理システムの開発を通じて、博物館における文化財情報の構造や管理運用に関する知見を蓄積していた。

所蔵品の情報は一般向けのためにインターネットでも様々な形で公開していたが、中でも国立文化財機構の4博物館(東京国立博物館、京都国立博物館、奈良国立博物館、九州国立博物館)が所蔵する国宝・重要文化財の高精細画像を5カ国語の解説とともに提供する「e 国宝」では、2010年のリニューアル時に標準的な検索 API (システム連携機能) である OpenSearch によるデータ出力機能を実装した。これを用いて開発し、翌年にリリースしたモバイルアプリ版「e 国宝」は大きな話題となり、広く一般の人々が文化財に親しむきっかけとなった。またこの OpenSearch により、当時の国立国会図書館デジタルアーカイブポータル「PORTA」(現在は国立国会図書館サーチに統合)との連携も実現した。

一方、Web上にあるデータの「意味」を表現する RDF (Resource Description Framework) 形式を用いることによって、出所の異なるデータを相互に関連付ける技術である Linked Data が様々な分野で活用されはじめていた。

国外では、大英博物館をはじめとして、すでに Linked Data によって所蔵品情報を提供し、問合せ言語である SPARQL を用いて検索可能としている館が存在していた。また欧州の文化財情報ポータルサイトである Europeana でも SPARQL で検索し、Linked Data によって情報を取り出すことが可能となっていた。しかし国内の博物館、美術館等では Linked Data や SPARQL の本格的な運用には至っていない状況であった。

2. 研究の目的

博物館の所蔵する文化財には極めて多様な分野のものが含まれており、他機関がもつ様々な情報とも複雑かつ有機的に関連しあっている。したがって、複数の機関の情報が互いに結びつき、それらを辿りながら情報を探索することができるような環境によって、文化財情報資源の価値は最大となると考えられる。現在の状況は、個々の機関によるデータベースの公開が進み、さらに複数機関のデータを集約した統合目録データベースあるいはポータルサイトが活用されている段階である。本研究はここからさらに進んで、統一的に検索できるだけでなく、Linked Data によって個々のデータが結びつき、意味のある関連性が表現され、より有効な文化財情報資源の探索と発見につながるような環境の実現を目指すものである。そのような環境では、ユーザはデータの網の目をたどるようにして文化財情報資源を探索することが可能になると考えられる。

3. 研究の方法

本研究では、Linked Data による文化財情報資源の探索と発見のための環境の構築に向けて、文化財情報を対象として具体的に Linked Data の技術を適用し、有効な情報探索の方法を探ろうとするものである。そのために、すでに公開されていた国立文化財機構の4博物館(東京国立博物館、京都国立博物館、奈良国立博物館、九州国立博物館)の所蔵品統合検索システム「ColBase」のデータを用いて、Linked Data によるデータ出力機能を開発することとした。

この開発を通じて、(1)具体的に利用可能なメタデータ語彙を様々な国際標準を中心とする既存の語彙から選定し、マッピングを行う。必要に応じて独自語彙の開発も検討する。(2)データベース内に人名・地名等の索引情報を整備し、国立国会図書館の典拠データ等の既存の典拠情報との関連付けを行う。(3)SPARQL Endpoint を実装し、様々な問い合わせを試行する。(4)既存の外部システムとの連携の可能性を検討する、といった検討や試行を行うことにより、目指す情報環境に必要なとされる技術の検証や手法の確立を試みることにした。

4. 研究成果

(1)メタデータ語彙の検討とマッピング

文化財情報の Linked Data 化のために用いるメタデータ語彙を検討するにあたって、まず国内外における類似機関での Linked Data による所蔵品情報の公開事例について調査を行った。それらの実態を踏まえた上で、既存の標準語彙を比較検討した結果、ICOM CIDOC が開発した

CIDOC CRM を本研究の基盤として採用することとした。さらに、ColBase 上のデータ項目を CIDOC CRM で表現し出力する実装の開発に向けて、具体的なマッピングを検討した。その結果、日本における指定文化財の区分（「国宝」「重要文化財」）や員数などについては、CIDOC CRM の語彙では不十分な場合があることがわかり、これらのデータのためには新たに独自の語彙を定義する必要があることが確認できた。

検討の結果として得られた CIDOC CRM および独自定義の語彙を「ColBase」データに適用する詳細については「ColBase データの RDF フォーマット案」として文書化した。また本フォーマット案は後述する SPARQL Endpoint の実装のための仕様の一部としても活用した。

（２）典拠データの整備

異なる博物館・美術館の所蔵品を相互に関連付ける情報の一つに、作者に関する情報がある。同じ作者による作品が複数の館に所蔵されていることは多く、有効に関連付けをすることができれば有用である。そこで、ColBase に含まれる作者の情報を抽出し、他機関のデータと関連付けを行うことができるよう、人名を索引化し、既存の典拠データとの関連付けを行った。

既存の典拠データとしては、日本人の人名の収録数が多い国立国会図書館典拠データと、美術作家の人名を国際的に収集している Getty Research Institute による Union List of Artist Names を用いた。いずれもすでに RDF 形式によるデータを公開しているため、それぞれの人物を表現する URI を ColBase の人名索引データに登録した。いずれの典拠データについても、ColBase の人名のうち一定程度が再録されているものの、再録されていない人名も多くあることがわかった。またそもそも ColBase 上で作者名とされている人名について、単に作品の署名等から転記されているだけでどのような人物であるか不明であるものが少なくないことや、分野によって作者名についての概念そのものが少しずつ異なることなども確認された。

（３）SPARQL Endpoint の実装

（１）で作成した既存語彙および独自語彙を用いたマッピングに基づいて、ColBase 上のデータを RDF として出力するプログラムを開発した。また出力した RDF ファイルを格納する RDF ストアを構築し、データを SPARQL クエリにより問い合わせることのできる SPARQL Endpoint を用意した。

（４）外部システムとの連携

当初は、（３）で構築した SPARQL Endpoint に対して様々なクエリを発行することによって、文化財情報の探索における Linked Data の有効性を検証する予定であったが、Linked Data は出所の異なるデータを連結することによって大きく活用が可能となるため、SPARQL Endpoint とは別に「RDF マッシュアップエンジン」を開発した。このエンジンでは DBpedia 等に代表される既存の SPARQL Endpoint を登録し、その Endpoint に対する問い合わせのテンプレートを追加しておく。そして、任意のクエリから得られた値をテンプレートの入力として用いることができるようにした。これにより、あるクエリの結果を用いて別の Endpoint から新たにデータを引き出してくるといった、連鎖的な検索が可能となった。これを用いることで、異なる機関から独立して提供されているデータを連結させて、文化財に関する情報資源を探索できることがわかった。

（５）今後の展望

上記（１）～（４）で得られた成果をもとに、美術史家や学芸員を始めとする専門家や、一般の人々が文化財情報資源を活用するためのシステムやアプリケーションの開発にむけて検討を進めたい。また ColBase 上のデータについての索引情報は現在非公開であるが、活用の可能性をさらに検討していきたい。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	宮崎 幹子 (Miyazaki Motoko) (50290929)	独立行政法人国立文化財機構奈良国立博物館・その他部局等・室長 (84603)	
研究分担者	田良島 哲 (Tarashima Satoshi) (60370996)	独立行政法人国立文化財機構東京国立博物館・学芸企画部・特任研究員 (82619)	
研究分担者	竹内 俊貴 (Takeuchi Toshiki) (70750149)	独立行政法人国立文化財機構九州国立博物館・学芸部文化財課・専門職 (87106)	
連携研究者	高野 明彦 (Takano Akihiko) (00333542)	国立情報学研究所・コンテンツ科学研究系・教授 (62615)	
連携研究者	後藤 真 (Gotou Makoto) (90507138)	国立歴史民俗博物館・研究部・准教授 (62501)	