

# 芸術・文化情報の Linked Open Data 普及に向けた現状と課題 – LODAC Museum を例に

嘉村哲郎<sup>1,2</sup> 加藤文彦<sup>3</sup> 松村冬子<sup>4</sup> 上田洋<sup>5</sup> 高橋徹<sup>5</sup> 大向一輝<sup>1,4</sup> 武田英明<sup>1,4</sup>

<sup>1</sup>総合研究大学院大学 複合科学研究科, <sup>2</sup>東京藝術大学 総合芸術アーカイブセンター/芸術情報センター,

<sup>3</sup>情報・システム研究機構 新領域融合研究センター, <sup>4</sup>国立情報学研究所, <sup>5</sup>株式会社 ATR-Promotions

昨年, 博物館情報を Linked Open Data で公開・共有できるシステム, LODAC Museum を公開した. 本稿では, LODAC Museum のメタデータ設計と情報の相互運用性について考察し, 従来の統合検索システムとの相違について論じる. また, 前年度からの更新として, 地理情報の追加と SPARQL Endpoint の利用, ならびに Semantic Media Wiki を用いた LOD アノテーションツールを紹介する. 最後に, 国内外の動向を取り上げ, 芸術・文化情報の LOD 普及に向けた現状と課題について述べる.

## LODAC Museum: The popularization of the cultural information resources as Linked Open Data

Tetsuro KAMURA<sup>1,2</sup>, Fumihiro KATO<sup>3</sup>, Fuyuko MATSUMURA<sup>4</sup>, Hiroshi UEDA<sup>5</sup>,  
Toru TAKAHASHI<sup>5</sup>, Ikki OHMUKAI<sup>1,4</sup>, Hideaki TAKEDA<sup>1,4</sup>

<sup>1</sup>The Graduate University of Advanced Studies (SOKENDAI), <sup>2</sup>Tokyo University of the Arts,

<sup>3</sup>Transdisciplinary Research Integration Center, <sup>4</sup>National Institute of Informatics, <sup>5</sup>ATR-Promotions Inc.

Last year, the LODAC project has been developed museum information sharing system called LODAC Museum. A LODAC Museum identifies and associates arts and cultural information resource from some different sources to provide integrated data that are published as Linked Open Data. In this paper, we try to give overview of our activities and current status of issues. Besides we discuss metadata interoperability of artistic and cultural fields in Japan.

### 1. はじめに

本研究は, 情報・システム研究機構新領域融合研究センターのプロジェクト, LODAC(Linked Open Data for ACademia)の一環として取り組んでいる. LODAC プロジェクトは, 2010 年度より開始し, 国内の学術に関する情報を Linked Open Data(LOD)と呼ばれる形式で公開・共有し, これらを社会全体で活用できる次世代ウェブの情報流通基盤構築を目指している. 2010 年度は, 博物館・美術館情報を LOD で公開・共有できるプロトタイプシステム, LODAC Museum を構築した[1]. 本システムは, 14 館の国公立博物館・美術館の資料情報と, 関連情報を 3 件収集して LOD で公開した. 2010 年度の主な取り組みには, ウェブサイトのデータ収集から, LOD で扱うためのデータ変換, 公開までの基本的な流れを構築し, 2011 年度はそのプロセスを再構成し, 具体的な LOD 活用に向けてアプリケーションを実装した. 本稿では, LODAC Museum を例に, 芸術・文化情報の LOD 普及に向けた現状と課題について述べる. 本稿の構成を示す. 次章ではプロトタイプシステムの概要とデータ形式に LOD を採用した理由. 3 章では現状のメタデータ設計と構造を示し, 情報統合と相互運用に合わせて類似研究のシステムと LODAC Museum の相違について述べる. 4 章では新規データの追加と共有・参照機能(SPARQL Endpoint)を利用したアプリケーションを紹介する. 5 章では国内外の芸術・文化に関する LOD 事例を取り上げ, 6 章では考察として, 芸術・文化情報の LOD 普及のために解決すべき主な課題と解決に向けた今後の展開について述べる.

### 2. LODAC Museum と RDF

#### 2.1. LODAC Museum

LODAC Museum は一般的な博物館・美術館等で使用されるコレクション管理システムとは異なる. 本システムは, LOD で情報を共有するための仕組みと活用方法・有効性を検証するプロトタイプであることから, コレクションを持つ組織が本システムをコレクション管理システムとして利用し, LOD で情報公開できる仕組みとしては作られていない. 各組織が保有する既存の情報を LOD で公開するためには別途データ変換・管理用のシステムを用意する必要がある.

LODAC プロジェクトでは, LOD で情報を扱うプロトタイプシステムを用い, 芸術・文化情報の LOD 普及に次の目標を掲げて研究を進めている.

1. ウェブに分散する同じ内容を含む複数の情報を LOD で扱い, 統合することで情報の相互運用性の向上を図る
2. 情報を共有することで, 多様な分野での利用を可能にし, 国内における情報流通の流動性と柔軟性を向上させる
3. 情報を公開する主体に対して何らかの還元が可能な次世代の情報流通基盤を構築する

#### 2.2. RDF によるデータ公開

RDF(Resource Description Framework)は, 1999 年に W3C から勧告された, Web で扱う情報を記述する共通の枠組みで, 情報(リソース)を主語(対象情報)・述語(情報項目/プロパティ)・目的語(項目値)の 3 つ組(トリプル)で表現するデータモデルである (図 1).

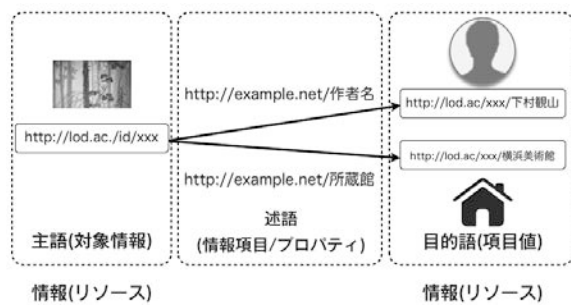


図1: 主語・述語・目的語で表現する RDF

RDF の特徴は、ウェブで広く情報共有・利用されることを念頭に設計されていることから、情報(リソース)の表現に URI(Uniform Resource Identifier) / IRI(Internationalized Resource Identifier) が用いられる。URI は、ウェブで情報源を特定できる唯一の識別子として標準化されており、とりわけ WWW では http から始まる URL アドレスとして用いられている[2]。一方の IRI は、識別子の表現に英数字のみを限定している URI に国際化文字対応させたもので、WWW では日本語ドメイン等で実用化されている[3]。RDF で情報を記述することは、URI や IRI の識別子を用いてそれぞれのリソースを定義することから、情報利用の際は、URI を指定することで一意にリソースを特定し、その内容を参照できる利点がある。LOD で公開する情報は、一般的に RDF データが用いられ、内容には独自に記述した情報項目と複数の RDF データのリンクで構成される。LOD は、RDF データの各項目に記述されるリンクを参照することで必要な情報を利用できることから、リンクでつながるデータ、すなわち Linked Data(LD)または Linked Open Data(LOD)と呼ばれる。これらは、ウェブに公開され、自由に利用できるオープンな定義[4]に適合した RDF データのこと。LODAC Museum では、収集した各資料情報を 1つのリソースとして扱い、それらを参照・引用して情報統合することから、各情報を一意に識別して管理できる必要があった。さらに、標準化された枠組みを用いることで、同様の形式で作られた異なる分野の情報と統合できる可能性があることから、情報流通の流動性と柔軟性の向上に適した方法として LOD を用いている。

### 3. 情報統合と相互運用

LOD によるアプローチは、情報(リソース)の要素に一般的なメタデータを用いた構造が定義できる点と、URI を用いた複数リソース間の関係性を定義できる点から、あらゆるリソースに対して柔軟性と流動性を与えられる利点がある。これは、ウェブに分散する多数の情報がある基準に従って統合する際の強力なツールになることから、その利用について国内外で非常に高く注目されている。本章では、LODAC Museum におけるデータ構造と情報統合について、類似研究事例を交えて論じる。

### 3.1. メタデータ設計

LODAC Museum で扱うデータは、情報源にウェブから収集した情報を使用している。これらの情報は、収集した情報源ごとにメタデータやデータ構造が異なるため、これらを LOD で扱うにはメタデータを再定義する必要があった。LODAC Museum のメタデータ設計では、作品や資料に単一のスキーマを用いて詳細な情報を記述するのではなく、できるだけ一般的に利用されているスキーマを用い、資料項目に対して適合するメタデータをマッピングしている。例えば、作品名や資料名を表すメタデータには dc:title と skos:prefLabel, 所有者については crm:P52\_has\_current\_owner 等、それぞれの項目にメタデータを再定義する。一つの情報に複数のメタデータをマッピングする理由は、利用者のメタデータに制限があるなど、メタデータに問い合わせができない場合に備え、複数のリクエストに対応できる設計にしている。現在、使用しているスキーマは DC, DC Terms, SKOS, FOAF, DCNLD, RDA,RDA2, CIDOC CRM, OWL, RDFS, GEO, ICAL, VCARD で構成され、メタデータについては各スキーマから必要な要素のみを参照し、要素に当てはまらない項目については独自のメタデータを定義している。

### 3.2. 情報統合

LODAC Museum では、典拠情報と統合情報の 2 種類の情報を扱う。前者は各情報源から収集したデータを RDF に変換した情報で、後者は筆者ら統合または新規作成した情報である。それぞれには典拠 ID, 統合 ID として一意の識別子を付与し、データベース上で管理する。典拠情報は、情報源から収集した内容をそのままデータにするため、資料タイトルや技法、制作年等を忠実に記述し、その内容に関する責任・権限については情報源に委ねることにしている。一方の統合情報は、多数の作品や作者の RDF データをリンクとして(すなわちリンクの統合)記述している。統合情報の内容は LODAC プロジェクトが統合・作成していることから本プロジェクトが責任を持つ。情報の統合・作成方法については参考文献を参照されたい[1]。

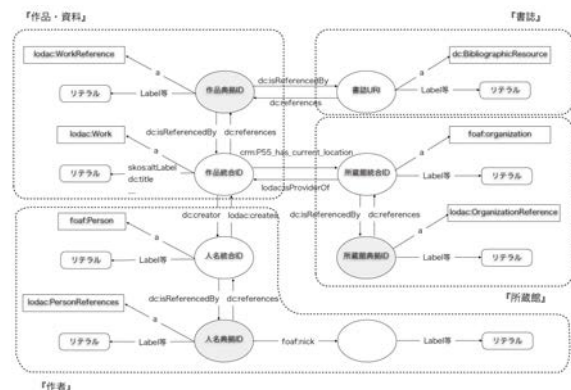


図2. LODAC Museum Data Structure

図2は LODAC Museum で扱う情報カテゴリを表す。カテゴリは 4 種類、作品・資料、作者・人物、施設、

書誌で構成される。書誌を除く各カテゴリには典拠情報と統合情報が含まれ、図中の白円は LODAC が内容を記述した統合情報、色付円は典拠情報を表す。長方形は情報に対するカテゴリを定義し、リテラルは各情報のメタデータに対する値が文字列として格納される。

現在、LODAC Museum が公開する統合情報は、作者・人名を起点とした作品・資料の名寄せと、所蔵館を起点とした作品・資料の名寄せ、作品・資料を起点とした所蔵館情報の名寄せに限られる。従って、これ以外を起点とする統合情報を利用とする場合は、利用者がそれぞれの視点で統合する必要がある。個々の情報に含まれる項目をデータとして利用する場合は、基本的には統合情報を参照し、そこに記述されるリンクを経由することで典拠情報内の各項目をデータとして利用できる。必要な項目のみデータとして取得したい場合は、その情報の RDF データを直接参照するか、筆者らが提供する SPARQL Endpoint と呼ばれる仕組みを利用する。このように、LODAC Museum では筆者らの視点で統合した資料情報とその典拠情報、関連情報が網羅的かつ柔軟な検索ができることに加え、必要な情報の必要な項目のみをデータとして利用できるシステムとして構築している。

複数のデータベースを統合し、横断的に検索できるシステム構築の研究に、人間文化研究機構の研究資源共有化システム[5]がある。このシステムは、“インターネット上に分散しているデータベースを、仮想的な単一データベースに見せる仕掛け”[6]を持つシステムで、機構が有する 6 機関のデータベースを統合し、1つのキーワードから関連する情報を横断的に検索できる仕組みを提供する。この研究では、LODAC Museum 同様の課題の 1 つにメタデータ問題について取り上げている。研究資源共有化システムでは、情報統合と検索を目的にするが、6 機関のデータベースが用いるメタデータが各々に異なるためそのままでは統合が困難であった。そのため、本研究が用いた手法は、各データベースから、意味的に共通するデータ項目を抽出し、それらの項目に対し、統合検索のためのメタデータを新たに定義している。この点では、筆者が収集した情報を RDF に変換する課程で適用するメタデータ適用方針と類似する。新たに定義されたメタデータは“共有化メタデータベース”[6]として専用のシステムに格納される。研究資源共有化システムのメタデータは独自のデータベース上、すなわち閉じられた環境にあることから、利用者がメタデータを参照する事や、メタデータ用いた情報の二次的利用が困難である。一方の LODAC Museum の手法は、すべてのメタデータは URI で参照可能なウェブにあることから、利用者はウェブを通じたメタデータ利用が可能で、独自の検索システムや閲覧システム等を作ることができる。研究資源共有化システムのように、特定のデータベースにメタデータが格納されている場合、そのシステムが提供する検索システムやインタフェースを利用する必要があり、利用者がデータベース上の情報を使って独自に開発することは困難である。その他に類似する統合検索システムには、京都大学の研究資源アーカイ

ブシステム[7]や国立国会図書館の PORTA[8]、自治体が提供する横断検索システム[9]などが存在する。

### 3.3. LOD と従来型統合検索システム

LOD で情報公開・共有を試みる LODAC Museum と先述した統合検索システムは、同じシステムと評されることがある。しかし、筆者らの目標の一つである共有という観点は、従来のシステムとは大きく異なる要素を持つ。LODAC Museum はそのインタフェースから、博物館・美術館資料の統合検索システムと捉えられるが、その本質は、データ標準化形式である RDF を用いて、必要な分だけ誰でも自由に情報を利用できるシステムである。例えば、従来の統合検索システムには API を用いて外部からデータ利用可能な仕組みを有する場合もあるが、API は提供するシステム単位で異なることから、利用者はその仕組みを理解する必要があり、手間を要する。一方の LOD では、ウェブ標準を用いていることから、一度仕組みを理解することで、LOD で公開される様々な情報を使うことができる。また、オープンな規格で情報公開されていることは、情報をどのように使うか、利用者任せられることができる。すなわち、情報検索した結果を閲覧するのみの従来型と、情報を自由に利用(閲覧・検索・システム構築・アプリケーションへの組み込み等)できる仕組みを有する LOD は大きく異なる。筆者ら博物館・美術館情報を統合(名寄せ)し、ブラウザで閲覧可能な仕組みやアプリケーションを公開していることは、将来的に様々な資料が各組織から LOD として公開された際の 1 つの活用・研究事例としての試みに過ぎない。LODAC Museum では、あらゆる利用者の視点で統合や作成された情報、あるいはそれらの情報を利用したアプリケーションが作られることを想定しており、このような情報流通の仕組みをつくることで、情報により流動性と柔軟性を与えることができる。

### 3.4. 相互運用性を考慮したデータ設計

相互運用性を考慮したメタデータ設計について、LODAC Museum では共通性の高いスキーマから必要な要素を参照してマッピングし、スキーマに要素がない場合は独自に要素を定義してきた。プロトタイプシステムを構築した当初、メタデータについては既存のあらゆるスキーマから要素を参照し、独自の定義は控える方針であったが、標準化メタデータユニバース[10]からも解るように、存在する標準化メタデータは膨大な数(2010年時点で105の標準化メタデータ)にのぼる。これらから必要な要素を探してマッピングすることは多くの時間を要し、非効率的であることから、LODAC Museum では先述する方法で定義することにした。しかし、この方法が専門的な情報資源を持つ組織の資料記述において、最適な選択とは言い難い。なぜならば、専門性や信頼性が高い情報は、公開された場合には典拠情報として利用される可能性がある。このような情報には、専門領域内で合意・策定されたメタデータを用いることが望ましい。大抵の標準化メタデータは特定領域、分野での利用を目的に作られてい

ることから、これらを使用することはその分野・領域内での相互運用性を高められることを意味する。ただし、標準化スキーマで記述出来ない情報がある場合は、メタデータを独自に定義することになる。その際は、使用するスキーマの記述規則に従い、定義することで、元のスキーマとの相互運用性を考慮した拡張が可能になる。相互運用性を尊重した既存スキーマの拡張利用については、メタデータ情報共有のためのガイドライン[11]でも触れられている。このような方法を用いた例に、大英博物館コレクションの LOD がある。2011年9月に公開されたコレクションデータベース LOD[12]には、メタデータスキーマに CIDOC-CRM が使用されており、スキーマで対応出来ない記述範囲については大英博物館独自の URI に CIDOC-CRM の規則に従ったメタデータ拡張がなされている。この方法は、使用するメタデータスキーマを必要限に押さえることでデータの冗長化を軽減できることから、管理や運用、設計面において有効的な手段の一つといえる。国内の事例では、2010年6月に国立国会図書館が公開した国立国会図書館ダブリンコアメタデータ記述 (DC-NDL) [13]がある。DC-NDL は、国会図書館の件名目録表にあわせて独自に定義され、Web NDL Authorities[14]の主要メタデータスキーマとして使用されている。DC-NDL では独自の定義のほか、標準化メタデータの DCMI Terms に対応する要素がある場合には関連付けて、上位概念に DCMI Terms が包含する形で設計されている。大英博物館は標準化メタデータに合わせて独自メタデータを拡張した方法を採用し、国立国会図書館は組織内の情報に特化した独自のメタデータ定義しつつ、それを標準化メタデータに合わせた方法と捉えられる。さらに、東京国立博物館の例では、保有するコレクションを管理するために専門性の高いデータモデル、ミュージアム資料情報構造化モデル [15]を独自に作成し、博物館運営に使用している。一方、現在の LODAC Museum では、様々な情報を独自の指針で収集・統合し、情報を再利用できるように公開(共有)していることから、情報に専門性を高めるのではなく、多数の領域・分野で広く利用されることを考慮して作成している。このような利用を想定とする情報には、一般的かつ共通性の高いメタデータを用いた記述をする事が好ましい。例えば、博物館資料において、タイトルを表すメタデータを考えた時、標準化規格の CIDOC-CRM を用いた場合は P102\_has\_title と記述し、一般的に使われるメタデータ、Dublin Core を適用した場合 dc:title になる。前者で記述した場合、博物館に特化したメタデータであることから、専門家はその意味や内容について理解できるが、一般的な利用者は意味を理解のために調べる必要がある等、手間を要する。対して、後者の場合は汎用性が高く、広く使われているメタデータであることから、その意味については容易に理解できる。メタデータを設計する際のスキーマ利用では、扱う情報の種類と用途の点から、次にあげる三つの方針が考えられる。

1. 限定範囲内、組織内利用重視:全て独自に定義したスキーマを用いる

2. 領域・組織間情報の相互運用、専門性重視:特定の分野・領域内で合意されているメタデータを用い、必要部分は独自に定義するスキーマを用いる
3. 一般性、利用効率性重視:全てにおいて汎用性の高い既存スキーマを用いる

1 については東京国立博物館が該当する。2 についてはより細分化でき、先述した大英博物館、国会図書館、LODAC Museum のそれぞれの設計方針に分ける事ができる。

## 4. RDF データの作成と共有

### 4.1. 地理・地名情報の LOD

本年度、新たに地理・地名情報の LOD を追加した。情報源には、住所に関する情報として「大字・町丁目レベル位置参照情報」と「郵便番号データ」、博物館などの公共施設の情報に「国土数値情報」、駅に関する情報に「駅データ」を利用した。これらの情報は、経緯度や所在地・施設名等がデータで公開されていることから、それぞれを RDF データに変換することで地理・地名情報を使用した情報統合ができるようになった。地名や住所情報を RDF で扱うために、都道府県市町村名、郵便番号、駅名を情報(リソース)として記述した。町丁目データには、「東京都 dc:hasPart 千代田区」のように地名に関係性を持たせることで、都→区→大→字→丁目の順になるよう定義している。経緯度情報については、地理情報の最小レベルである町丁目情報に記述される(図.3)。その他の郵便番号や駅名、公共施設については、それぞれに含まれる各種名称や住所、経緯度情報を用いて情報が関連付けされる。

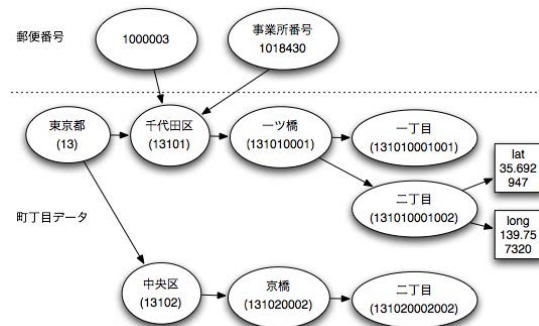


図3. 町丁目データのRDF階層図

地理・地名に関連する情報を LOD にすることで、住所を基準にした情報検索や経緯度情報から関係を辿り、関係の上位に含まれる情報を抽出することが可能になった。しかし、市町村合併による地名変更や頻繁に更新される郵便番号など、随時対応が必要な情報の更新については、大きな課題として残されている。

### 4.2. 博物館・美術館資料情報の追加

2010年度に公開した LODAC Museum は、主に14館の資料情報と関連情報3件を情報源としている。本年度はこれを大幅に増加し、計80館からスクレイピングによるデータ収集を試みた。今回収集した情報に

は美術系、図書系、文書系、自然系、考古系等いわゆる博物館法で定義される各形態の資料が含まれるが、実際にシステムに追加した情報は、既存の LOD と統合・関連付け可能性のある美術系資料を中心とした情報源を選択した。さらに、その中でも構造が簡潔で RDF 変換が容易と思われる美術系 39 館から計 60,629 件の作品情報を抽出した。今回、データの収集範囲が拡大したことから、データ管理方法を見直し、RDF データ作成までの流れをシステム化した(図 4)。

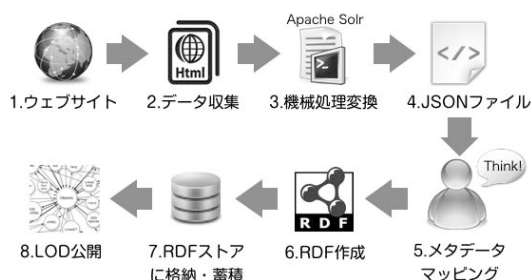


図 4. RDF データ作成のながれ

今回、各資料の HTML 管理及び項目抽出に Apache Solr を用いたデータ管理・変換システムを構築した。これにより、スクレイピングルールを用いたウェブからの HTML データ収集と機械処理可能な JSON データを生成できるようになった。JSON データは資料ごとに生成され、内容にはスクレイピング時に付与するシステム管理情報(タイムスタンプや MD5 等)と実際の資料情報が含まれる。さらに、ここから RDF データを生成するためには、変換時に使用するメタデータマッピングテーブルを作成する。テーブルには、各 JSON データの値を目視で確認し、その値に適用できるメタデータをテーブルに記述する。今回実施した方法では、スクレイピングした 80 館それぞれの JSON データ項目を表に並べ、それらを LODAC Museum で使用するメタデータにマッピングした。JSON データはスクレイピングした館やコレクション単位で内容が異なるため、それぞれに対してマッピングテーブルを作成する必要がある。その後、テーブルを利用して 1 つの JSON データごとに RDF データを生成し、最終的にデータストアに格納することで LOD として公開する。今回のシステム化では JSON を RDF に変換する際に失敗するデータが多数あった。その原因の 1 つに、コレクションや館ごとにマッピングしているメタデータテーブルが、一部のデータに対して適用できなかったことがあげられる。例えば、美術館 A から収集したデータ 1000 件のうち 800 件は同じデータ構造であったが、処理途中から異なる構造であることが判明し、館やコレクション単位で作成したマッピングテーブルが対応できなかった。その他に、収集段階において意図しないデータが含まれる等、例外的な対応が多く、前回と同様にウェブから収集した情報の RDF 作成には非常に多くの時間を要した。

#### 4.3. RDF アノテーションツールとしての Semantic Media Wiki の利用

LODAC プロジェクトでは、LODAC Museum の公開に伴い、将来的に誰もが RDF データ内容の補完や作成ができるように、アノテーションツールの開発を試みた。基盤にしたツールは、情報を RDF で出力する仕組みを標準で備える Semantic Media Wiki を使用し、部分的に改良を加えることで LODAC Museum と連携可能かテストした。しかしながら、Semantic Media Wiki には、大きな問題が二つあること判明した。一つは、Wiki の仕様により、全角日本語文字を使用した項目を RDF で出力した時、RDF に記述される URI が自動的に独自形式に変換されるエンコーディングの問題がある。具体的には、URI に日本語が含まれていた場合、本来ならば下記 A で出力されるべきところ、B のように出力される。

A. <http://lod.ac/wiki/%E5%A6%82%E6%9D%A5>

B. <http://lod.ac/wiki/-E5-A6-82-E6-9D-A5>

この場合、実際にリンク先を参照するとエラーになることから、Wiki が出力する日本語を含めた URI については「%」が「-」に置き換わることがわかった。そこで、Wiki 側で処理される RDF データに対して、データとして表示する際に、本来の形に変換する処理を加えることで簡易的に問題を回避している。二つ目は、Wiki から出力する RDF データはそのまま LOD で利用できない点がある。そこで、これらを LODAC Museum で利用できるようにするため、RDF で出力するデータに対しては URI で参照できるように処理を加え、LODAC Museum の各リソース表示画面に Wiki 専用のリソースタブを追加することで実現している。このタブは Wiki 上の情報有無に関係なく、LODAC の ID を元に自動的に生成されている。また、URI については、Wiki の情報と各情報源から収集した情報の区別ができるように異なる URI で管理している。

次に Wiki を用いた情報作成について説明する。情報を記述するためには、初めに 4 つの作業が必要になる。1.メタデータ(プロパティ)作成、2.データテンプレート作成、3.データ入力用フォーム作成、4.データ入力。一度 1~3 を実施することで以後データ入力のみとなる。最初のメタデータ(プロパティ)作成では、入力する情報に付与するメタデータの定義とデータの型を指定する(図 5)。

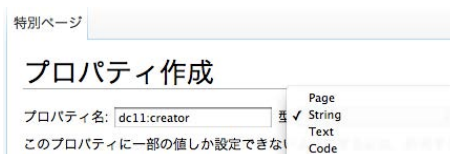


図 5. メタデータ(プロパティ)作成

その後、作成する情報に必要なメタデータを指定したテンプレートを作成する。テンプレートには、1 つのフィールドにつき 1 つのメタデータを指定する。メタデータについてはプルダウンメニューから選択できるが、登録されている全てのメタデータ表示されることから、大量にメタデータを作成した場合は指定に手間がかかるなどの欠点がある(図 6)。



図 6. メタデータテンプレートの作成

テンプレート作成後は、情報を入力するためのフォームを作成する。フォームには複数のメタデータテンプレートを組み合わせることができる。フォーム作成後は実際にデータを入力する。入力にはテンプレートで指定したメタデータセットに従い、テキストボックスへの入力やプルダウンでの選択が可能になる(図 7)。

仏像フォームを作成：八宗論大日如来像

図 7. データ入力フォーム

作成されたデータは Wiki のページと LODAC Museum のタブから参照できる(図 8)。ただし、Wiki のデータはタブに直接表示するのみで、LODAC Museum の全文検索や SPARQL Endpoint からの直接利用は出来ない。現在のところ、Wiki のデータは Wiki 専用の RDF ストアに格納されることから、SPARQL Endpoint を利用したい場合は Wiki で用意したものを使う必要がある。



図 8. LODAC Museum の Semantic Media Wiki 情報

このように、Semantic Media Wiki を用いた RDF データの作成・編集ができるようになった。しかし、RDF データの管理やメタデータの容易な選択方法等、細かな問題が多数存在する。また、テンプレートやフォームの作成・編集には Wiki や RDF に関する知識が必要になることから、利用には多少の訓練が必要になる。現状、プログラム上の変更や処理機能の追加など多くの課題を抱えていることから、Semantic Media Wiki をアノテーションツールとして改良を継続するか、別の方法で実現するか検討する必要がある。

4.4. SPARQL Endpoint の利用

SPARQL Endpoint とは、RDF に特化した問い合わせ言語を用いた、検索のためのインタフェースである。今年度は、本機能の実装により、LODAC Museum の情報を柔軟に検索することや、外部プログラムが RDF データに対して直接問い合わせできるようになった。SPARQL の利用は、従来の全文検索システムと異なり、必要とするデータのみを抽出することや、抽出するデータを統合して結果出力ができる等、情報を柔軟に扱えることが大きな特徴にある。SPARQL を利用したウェブアプリケーション実装に、ATR-Promotions 上田氏の開発による Twitter 連動アプリケーション「ミュージアムへ行こう！」[16]がある。このアプリケーションは、Twitter 利用者のつぶやきに経緯度の情報が含まれているか、内容に地名が含まれていた場合に反応し、利用者の最も近くにあるミュージアムを自動的に検索してその概要やルート情報を応答する(図 9)。



図 9. go2museum 応答例

これらの情報は SPARQL を用いて LODAC Museum 上のデータに問い合わせることで取得している。

LODAC Museum では、プログラムからの利用に加えウェブブラウザからクエリを実行できる SPARQL 検索フォームを用意し、実際に試することができる。SPARQL 利用する事で容易に用語の種類を調査できる例を示す。博物館・美術館資料においては度々表記揺れや使用する用語種類が問題として取り上げられることがある。下図は LODAC Museum の全作品情報から技法に“リトグラフ”が含まれる表記を抽出した。label は作品・資料に記述されている内容、countName はその記述が使われている作品数を表している。

label	countName
リトグラフ(単色) @ja	91
リトグラフ、オフセット・ポスター・1 @ja	54
リトグラフ、ドライポイント、紙 @ja	40
カラー・リトグラフ @ja	35
カラーリトグラフ @ja	33
カラーリトグラフ、紙 @ja	30
リトグラフ、手漉紙 @ja	30

図 10. SPARQL を用いた情報抽出例

その結果、LODAC Museum 上では 282 種類のリトグラフに関する表記があることが判明した。このように、技法や材質などの表記や種類、頻度について統計的な調査が容易に実施できる。SPARQL の利用は、使い次第で基礎研究的な利用からウェブアプリケーションへの組み込みなど、新しい情報活用への期待が非常

に大きい。SPARQL を用いたその他のサンプルクエリは LODAC サイトを参照されたい[17]。

## 5. 芸術・文化情報における LOD 事例

昨年から今年にかけて、芸術・文化情報に関連する LOD の取組みが活発化している。本章ではいくつかの事例を紹介する。3 章で取り上げた大英博物館では、英国内の博物館と情報系研究者による博物館情報利用の活発な議論の末、LOD 公開に至っている[18]。英国では BBC 放送が世界に先駆けて、Linked Data を利用したウェブサイトを開示していることから、これらの先行事例も後押ししたものと考えられる。大英博物館コレクション LOD では、LODAC Museum 同様に SPARQL Endpoint を公開し、自由にコレクション情報を使うことができる。

EU 圏においては博物館や図書館、文書館のデジタルデータを大規模に公開する Europeana[19]がある。Europeana では、2011 年 9 月にメタデータの著作権制限にクリエイティブコモンズ・ライセンス CC0 を適用する協定内容を発表した [20]。これにより、将来的には公開する全資料情報に対して CC0 を適用すること、すなわちパブリック・ドメインや LOD での公開・利用を目指している。しかし、適用には EU 各国情報源の同意が必要であることから、しばらく時間を要すると見られる。ただし、LOD については試験的に公開されており、現在はアムステルダム国立博物館 70,000 点の資料情報が公開されている[21]。これらには用語等の典拠情報に Getty の ULAN や AAT[22]を拡張した AATNed[23]、関連情報に DBPedia、地理情報に GeoNames を利用する。この 1, 2 年間は、文化・芸術情報を LOD で公開する試みが多く見られ、データに関するだけでなく、問題に対する取り組みも見られる。例えば、デジタルデータをウェブで公開する場合、その情報をどの程度自由に扱えるようにするのか、利用範囲に関する問題が度々取り上げられる。この問題について、2011 年 6 月、サンフランシスコで行われた図書館、文書館、博物館の LOD を考える国際会議[24]で 1 つの方針が示された。“*a 4-star classification-scheme for linked open cultural metadata*”としてクリエイティブコモンズおよびオープンデータコモンズのライセンスに沿った適用の推奨を示している。具体的には各情報に対して CC0,CC-BY,CC-BY-SA を表記する。クリエイティブコモンズ・ライセンスの表記を用いた博物館情報の公開事例に、2011 年 10 月、米国のウォルターズ美術館が約 1 万点の資料情報、画像データに対して CC-BY-NC-SA を適用した例がある [25]。さらに同時期、国際博物館会議 ICOM からは、博物館情報に対する LOD 設計指針として“*ICOM Recommendations on Linked Open Data for museums*”[26]が発表された。内容には、博物館情報を LOD として設計する際の必要事項や留意点が記述される。このように、欧米では芸術・文化情報のオープン化や LOD 取り組みが非常に活発化している。国内の事例では、2010 年より開始した横浜 LOD プロジェクト[27]があげられる。本プロジェクトは地域情報資源のオープン

データを目指しており、横浜市芸術文化振興財団や横浜観光コンベンションビューローが持つ芸術や観光情報を LOD による公開を始めている。さらに、国内の芸術・文化情報基盤整備の一環として、東京藝術大学総合芸術アーカイブセンターでは、文化財保存学日本画研究室編日本画用語辞典[28]を元にした用語の LOD を進めている。主に文字情報として公開される情報には、クリエイティブコモンズ・ライセンスの CC-BY-NC-ND が付与される。このライセンスはクレジット表示、非営利目的、改変禁止で、先述した LOD 方針よりも制限が厳しいものになるが、この制限を適用した RDF データを公開することで、その需要を図るねらいがある。このように、国内外を含めて少しずつではあるが、芸術・文化情報を LOD として公開する動きが高まりつつある。

## 6. 考察

様々な LOD に関する動きが活発化する中で、課題も多い現状にある。本章では、考察として LOD 普及に伴う主な課題とその解決への展望について述べる。

データ構造及びメタデータ設計の課題では、データが階層的構造や互いに関係性を持つ場合に、RDF の仕様に関する知識と領域・分野に関する知識が必要になることから、これらを理解できる人的リソースが必要になる。メタデータについても同様に、適用するメタデータの選択や独自定義の際の設計ができる知識が必要になる。しかしながら、文化・芸術系では情報技術に関する知識を有する人物が少ない。この分野では情報に関する知識を持つ人材の育成を視野に入ると共に、データ作成時の負担を軽減するためになんらかの支援が必要である。例えば、特定の領域・分野で使える RDF アノテーションツールや、ウェブに存在する既存のメタデータを容易に把握するための支援ツールなどが考えられる。既存の情報を利用して LOD に参加するには、データベース等の情報を RDF 化する方法やツールが必要になる。情報が単純な CSV データ等の場合は、irON[29]や Google-Refine[30]等のツールを用いて変換することができる。リレーショナルデータベース(RDB)の情報を RDF データに変換する場合は、W3C が仕様の策定を進めている変換言語、R2RML[31]を用いる方法がある。R2RML を使用した変換ツールには、ドラフト仕様をベースにした Perl ライブラリ[32]が利用できる。また、RDB からの RDF データ作成方法では、RDB のデータをダンプや CSV 等で出力した後、別途 RDF に変換する手法も考えられる。この方法を用いた場合は、RDB のデータと RDF データを個別に管理出来ることから、既存のシステムに手を加えることなく RDF を管理できる利点がある。さらに、非公開情報がある場合には、変換時に対象から外せる等、その利用範囲を柔軟に決められることから、変換手段の一つとして検討できる。

LOD 利用に関する問題では、日本の芸術・文化領域において自由に利用できる典拠となるようなシソーラスが少ないことから、シソーラスを基準にした検索や情報統合が難しい現状がある。例えば、人間文化研

究機構の研究資源共有化システムが持つ時空間情報が LOD 化されたならば、時間軸に関する典拠情報ができることから、時代や年代を基準にした統合や横断検索が可能になる。用語の問題についても、何らかの典拠情報が LOD で公開されることで、美術館や博物館の資料記述に利用できることが考えられる。これは、図 10 に示すような 1 つの技法に対して 300 近い表記を減らせることも考えられる。LOD による情報公開は、従来のデータベースシステムでは困難であった、自由な形で情報を取り出し、その情報に新たな創造を加えて共有・利用できる新しい仕組みを実現する。これは、公開側・利用側の双方に新しい価値を創出できる次世代の情報流通基盤といえる。本研究では芸術・文化情報における LOD の普及と、分野の活性化を目指して現在ある課題に取り組んでいく。今後の LODAC Museum では、データ構造に関する議論の継続と、スクレイピングした残り 41 館のデータ追加に関する検討、ならびに ULAN や AAT 等の既存シソーラスによる情報統合を試みる予定である。

## 謝辞

本研究を進めるにあたり、LOD に関する活発な議論に参加できたことを LODAC Project メンバーに感謝致します。また、文化情報の取り扱いについて様々な意見を頂いた Linked Open Data Japan[33]の皆様方に敬意を表します。

## 参考文献

- [1] 嘉村哲郎,加藤文彦,上田洋,高橋徹,大向一輝,武田英明:Linked Open Data による多様なミュージアム情報の統合,人文科学とコンピュータシンポジウム 2010 論文集,2010.
- [2] T.Berners-Lee, et al.,Uniform Resource Identifiers (URI):Generic Syntax, 2005.1,Standards Track, <http://labs.apache.org/webarch/uri/rfc/rfc3986.html>
- [3] M. Duerst and M. Suignard, Internationalized Resource Identifiers (IRIs), 2005.1, Standards Track, <http://www.ietf.org/rfc/rfc3987.txt>
- [4] Open Knowledge Foundation, <http://opendefinition.org/okd/japanese/>
- [5] 人間文化研究機構研究資源共有化システム, <http://www.nihu.jp/sougou/kyoyuka/index.html>
- [6] 原正一郎:時空間情報処理ツールの研究・開発,情報処理学会研究報告,Vol.2009-CH83,No.7,2009
- [7] 京都大学研究資源アーカイブ, <http://www.rra.museum.kyoto-u.ac.jp/>
- [8] 国立国会図書館デジタルアーカイブポータル (PORTA) , <http://porta.ndl.go.jp/>
- [9] しまねバーチャルミュージアム, <http://www.v-museum.pref.shimane.jp/>
- [10] Seeing Standards:A Visualization of the Metadata Universe, <http://www.dlib.indiana.edu/~jenrile/metadatamap/>
- [11] メタデータ情報共有のためのガイドライン, <http://www.meta-proj.jp/A03.pdf>
- [12] British Museum Semantic Web Collection Online, 2011.9.5, <http://collection.britishmuseum.org/>
- [13] 国立国会図書館ダブリンコアメタデータ記述, <http://www.ndl.go.jp/library/data/meta/2010/06/dcdndl.pdf>
- [14] Web NDL Authority, <http://id.ndl.go.jp/auth/ndla>
- [15] ミュージアム資料情報構造化モデル,東京国立博物館, <http://webarchives.tnm.jp/docs/informatics/smmoi/>
- [16] ミュージアムへいこう @go2museum, <http://twitter.com/#!/go2museum>
- [17] LODAC, <http://lod.ac/wiki/SPARQLスニペット>
- [18] Glaser, H., Museum data where next – consuming linked data. In: Cultural Heritage and the Semantic Web, 2011.01.13, British Museum. , <http://eprints.ecs.soton.ac.uk/21894/>
- [19] Europeana, <http://www.europeana.eu/portal/>
- [20] European Public Domain Charter, <http://www.version1.europeana.eu/web/europeana-project/public-domain-charter-en>
- [21] ThoughtLab: Think With Us Linked Open Data, [http://www.europeana.eu/portal/thoughtlab\\_linkedopendata.html](http://www.europeana.eu/portal/thoughtlab_linkedopendata.html)
- [22] Getty Vocabularies-The Union List of Artist Names, <http://www.getty.edu/research/tools/vocabularies/>
- [23] Nederlandstalig Art & Architecture Thesaurus, <http://www.aat-ned.nl/>
- [24] International Linked Open Data in Libraries Archives and Museums Summit (LOD-LAM), 2011.6.6, <http://lod-lam.net/summit/2011/06/06/proposed-a-4-star-classification-scheme-for-linked-open-cultural-metadata/>
- [25] ウォルターズ美術館プレスリリース,2011-10-4, [http://thewalters.org/news/releases/pressdetail.aspx?e\\_id=298](http://thewalters.org/news/releases/pressdetail.aspx?e_id=298)
- [26] ICOM Recommendations on Linked Open Data for museums, <http://www.collectionslink.org.uk/discover/sustaining-digital/1082-cidoc-recommendations-for-object-identifiers-for-linked-open-data>
- [27] 横浜 LOD プロジェクト, Open Community Data Initiative (OCDI) ,<http://ocdi.jp>
- [28] 東京藝術大学大学院文化財保存学日本画研究室,2007,「図解 日本画用語事典」,東京藝術大学,221pp
- [29] OpenStruct:irON, <http://openstructs.org/iron>
- [30] Google-refine, <http://code.google.com/p/google-refine/>
- [31] R2RML: RDB to RDF Mapping Language, <http://www.w3.org/TR/r2rml/>
- [32] RDF::RDB2RDF::R2RML,<https://metacpan.org/module/RDF::RDB2RDF::R2RML>
- [33] Linked Open Data Japan, <http://www.facebook.com/groups/130625107020532>