

# LODAC Museum: Linked Open Data による博物館情報の統合と活用

## LODAC Museum: Integration and Utilization of Linked Open Data in Museums

松村冬子\*1      嘉村哲郎\*2\*3      加藤文彦\*4      小林巖生\*5      高橋徹\*6      上田洋\*6  
 Fuyuko Matsumura      Tetsuro Kamura      Fumihiro Kato      Iwao Kobayashi      Toru Takahashi      Hiroshi Ueda

大向一輝\*1\*2      武田英明\*1\*2  
 Ikki Ohmukai      Hideaki Takeda

\*1国立情報学研究所      \*2総合研究大学院大学      \*3東京藝術大学  
 National Institute of Informatics      The Graduate University for Advanced Studies      Tokyo University of the Arts

\*4情報・システム研究機構      \*5Open Community Data Initiative  
 Transdisciplinary Research Integration Center

\*6株式会社 ATR-Promotions  
 ATR-Promotions, Inc.

This paper reports the progress of LODAC Museum project which aims to generate Linked Open Data of museum information in Japan. Collection information of 53 museums is scraped from their websites and converted to LOD. In order to efficiently collect massive amount of data and convert it to LOD, LODAC Distiller, a LOD generator is developed and it consists from two main parts; data processing and metadata design. Semantic MediaWiki is also introduced to enable user annotation to LOD based on metadata schema defined by them. Moreover, three applications that use LODAC Museums and other LODs are developed and discussed; Yokohama Art Spot, CamCat and go2museum.

### 1. はじめに

近年、Web 上でデータを公開、共有、再利用する仕組みである Linked Open Data (LOD)[Bizer 2009] が注目されている。特に欧米を中心として政府、新聞社や放送局などのマスメディア、図書館や博物館、研究機関など数多くの主体が LOD を公開しており、互いにリンクを結び合いつつある。

我々は 2010 年 4 月より、国内において広く学術に関連する情報を収集し LOD として公開する LODAC(Linked Open Data for ACademia) プロジェクト [武田 2011]\*1 を開始した。本プロジェクトは、情報・システム研究機構 新領域融合研究センターにおける異分野研究資源共有・協働基盤の構築を目的としたプロジェクトのサブプロジェクトであり、現在は博物館情報をはじめ、生物多様性情報、地理情報など、様々な学術情報および公共情報を LOD として公開している。

このうち、博物館情報の LOD である LODAC Museum は、国内の博物館におけるコレクション情報を収集し、LOD 化したものである。昨年度は、複数の博物館の Web サイトから収集した情報、地理情報、横浜におけるアート関連のイベント情報などの LOD 化、それらのデータの検索を実現する SPARQL Endpoint の設置について報告した。本稿では、より多くの博物館から新たに収集したデータの統合、Semantic MediaWiki を用いた LOD へのアノテーション機能の実装、および LODAC Museum のデータを用いて開発したアプリケーションについて報告する。

### 2. Linked Open Data (LOD)

LOD とは、個々の事物を Web 上で一意な URI で指し示すことのできるリソースとして、その関係をリンクで表現することで、データの公開、共有、再利用を促進するための仕組みである。LOD によって互いに繋がったデータ群は、データの Web (Web of Data) と呼ばれるグローバルなデータ空間を構成している。LOD は、1) あらゆる事象に対して URI を付与する、2) HTTP 経由で URI を参照可能とする、3) URI を参照した際は情報が閲覧可能とする、もしくはデータは RDF や SPARQL など標準化された技術でアクセスできるようにする、4) 他の URI へのリンクを含める、という 4 つの原則に基づいている。これより、各データは URI によって Web 上で一意的に指し示され、HTTP や RDF など標準的な仕様に基づいてアクセスすることが可能になることから、オープンデータ化の手段として利用できる。また、LOD では関連するデータ同士をリンクで結び合うことで情報を表現することから、一箇所に集約するのではなく、分散しているデータをリンクすることで統合的に扱うことができるため、分野内でのデータ共有の手段としても有効である。さらに、分野にかかわらず標準化された一定の形式による記述や検索が可能であること、オントロジーとして表現されるスキーマを理解して利用することが可能であること、分野を横断したリンクの結合が容易であることにより、分野を超えたデータ共有にも有効である。

### 3. LODAC Museum

#### 3.1 博物館情報の現状

現在、国内には 6,000 館近くの博物館があり、その一部が Web サイトを公開し、コレクション情報などを提供している。しかし、これらの Web サイトは各博物館が独自に構築しており、それぞれ異なるフォーマットで公開されている場合が多い

連絡先: 松村冬子, 国立情報学研究所 情報学プリンシプル研究系, 〒101-8430 東京都千代田区一ツ橋 2-1-2 学術総合センター, 03-4212-2666, fuyuko@nii.ac.jp

\*1 <http://lod.ac/>

こと、HTML や PDF など公開されている場合が多いことから機械的に取得し、処理することは容易ではない。また、博物館間でのデータの連携がないため、たとえば同じ作者に関する記述が異なる館の Web サイトに存在した場合でも、それが同一の作者に関する言及である、という事実を機械的に認識することは難しい。そのため、ある作者の作品が国内もしくは国外のどこにいくつ所蔵されているのかなどを横断的に調査することも不可能である。文化遺産オンライン<sup>\*2</sup>では、参加館から収集した情報の統合検索を可能にしているが、作品、作者などについて Web 上で一意に参照し、認識を共有するための典拠のような役割は果たしていない。このような背景から、LODAC Museum では、国内の様々な博物館のコレクション情報、そして DBPedia (Wikipedia の LOD) から、作品、作者、所蔵館などの情報とそれらの関係を Web 上で一意に指し示せるように LOD 化し、多様な情報源から博物館情報を統合的に得られる仕組みを構築することを目指している。本来はデータを所有している博物館が LOD を公開し、それらをリンクによって統合することが望ましいが、現在は本プロジェクトで博物館の Web サイトをスクレイピングし、LOD 化を図っている。博物館分野の LOD としては、ヨーロッパにおける図書館、公文書館、美術館などから収集したデータを統合的に扱える Europeana[Haslhofer 2011]<sup>\*3</sup>と呼ばれるサイトが有名である。Europeana においても、LOD によるデータの公開が行われており、LODAC Museum とこれらの海外の博物館情報の LOD をリンクさせていくことで、国境を超えて横断的に博物館情報を利用することが可能になる。

### 3.2 データ表現と統合

LODAC Museum では様々な博物館の Web サイトから収集した情報を用いているが、情報源ごとにメタデータが異なることから、LODAC Museum では新たにメタデータを定義した。他の LOD と連携して活用される場合などを想定して、できるだけ一般的に用いられているメタデータを用い、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た。なお、現在使用している既存のメタデータスキーマは、DC, DC Terms, SKOS, FOAF, DCNDI, RDA, RDA2, CIDOC CRM, OWL, RDFS, GEO, ICAL, VCARD などである。

LODAC Museum では、図 1 に示すように、各情報源から得られた作者、作品、所蔵館、関連資料にはそれぞれに情報源ごとの ID が付与される。ある作者に関する情報が複数の情報源から得られる場合は、複数の情報源ごとの ID に加えて共通の統合 ID が付与される。この作者について言及する場合は統合 ID を用い、詳細な情報はその情報源となる ID をリンクをたどって参照することでアクセス可能である。これにより、情報源ごとに持つデータと、統合されたデータをそれぞれ独立させることができる。なお、LODAC Museum では、このような作者情報の統合を行う際に、基盤となる情報として日本美術シソーラス [福田 1997] を参照した。

### 3.3 コレクション情報の増補

昨年度の人工知能学会全国大会においては、15 館の資料情報と 3 件の関連情報を情報源として構築した LODAC Museum について報告した [武田 2011]。今年度は新たに計 80 館の Web サイトからデータ収集を行い、そのうち既存のデータおよびメタデータと関連性の高い美術系資料が主となる 38 館から計 60,629 件のデータを新たに追加した。今回のデータ増補にお

いては、大量のスクレイピングを行うことから、図 2 のように RDF データの生成をシステム化した [加藤 2011]。

まず、Web サイトをスクレイピングしたデータから、Apache Solr を用いて「作者：藤田嗣治」などの記述から key と value のペアを抽出し、それを JSON の中間ファイルとして出力した。次に、各館の Web サイトから収集された JSON に記述されている key の内容と value の値から、手作業で key に適切なメタデータ語彙を割り当て、マッピングルールを作成する。最後に、作成されたマッピングルールに基づいて各 JSON ファイルから RDF を自動生成し、RDF ストアに格納することで、これらのデータを LODAC Museum に追加する。これにより、データ処理とメタデータの設計を分離することができ、効率的な RDF の生成が可能になる。

## 4. Semantic MediaWiki を用いたユーザによるアノテーション

現在の LODAC Museum では、博物館の Web サイトをスクレイピングして取得したコレクション情報や、DBPedia のデータを統合して LOD 化している。しかし、それらの情報は博物館の学芸員や Wikipedia を編集したユーザなどの視点に偏っており、異なる視点により情報を追記したいという要望を持ったユーザが現れる可能性がある。そこで、Semantic MediaWiki [Krötzsch 2006] と呼ばれる Wiki を利用して、ユーザが入力フォームから LODAC Museum にデータを追加できるように実装を行った [嘉村 2011]。

ここでは、統合 ID4301 (<http://lod.ac/id/4301>) の大日如来像のデータに対して、ユーザが新たに仏像にまつわるメタデータを記入したい場合を例として、Semantic MediaWiki によって LODAC Museum にデータを追記する方法を説明する。まず、ユーザは Semantic MediaWiki で仏像に関するメタデータスキーマ、つまり項目名と対応する語彙の組み合わせを作成する。このメタデータスキーマには、広く一般に使われている既存の語彙でも、独自の語彙でも、ユーザが適していると判断した語彙を用いることができる。独自の語彙は Semantic MediaWiki で定義することが可能である。次に、データを追記したい LODAC Museum のデータの統合 ID をページ名として Semantic MediaWiki において新しいページを作成し、図 3 のように作成した仏像に関するメタデータスキーマに基づいて表示される入力フォームに値を入力し、保存する。一

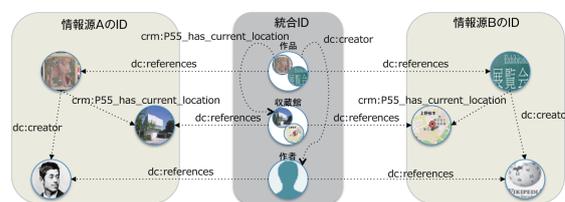


図 1: LODAC Museum のデータモデル



図 2: LODAC Museum のデータ生成の流れ

\*2 <http://bunka.nii.ac.jp/>

\*3 <http://europeana.eu/>



図 3: Semantic MediaWiki によるデータへのアノテーション

方, LODAC Museum の統合 ID4301 のページでは, 図 3 に示すように, Semantic MediaWiki において保存された統合 ID4301 に対応するデータを新たな情報源として 1 つのタブに表示する. これにより, 博物館のコレクション情報から構築した LODAC Museum のデータは保持したまま, 新たにユーザにより追加されたアノテーションを同時に提供することが可能となった. また, 仮に相反する内容の複数の情報源がある場合でも, 情報源ごとのデータを明確に分離して ID を付与し統合 ID にリンクさせることで, それぞれの主張を共存させることができることから, 専門家によるアノテーションにも利用できる.

## 5. アプリケーションの開発

本節では, LODAC Museum のデータを用いて開発された「Yokohama Art Spot」, 「ふおとぶらり」, 「ミュージアムへ行こう!」の 3 つのアプリケーションについて紹介する.

### 5.1 Yokohama Art Spot

2. 節でも触れたように, LOD によるデータ公開のメリットには, 分野内でのデータ共有だけでなく, 分野を超えたデータ共有の促進が含まれる. Yokohama Art Spot は, 本プロジェクトで構築している国内の博物館のコレクション情報の LOD である LODAC Museum, 公益財団法人横浜市芸術文化振興財団が構築しているアートに関する横浜の地域情報の LOD である「ヨコハマ・アート LOD」, エヌ・ティ・ティレゾナント株式会社が運営している位置情報 Q&A サービスの「PinQA」\*4 の LOD, という 3 つの異なる LOD を連携活用した横浜市内のアート関連情報を提供する Web アプリケーションである [松村 2011].

Yokohama Art Spot の構成を図 4 に示す. ユーザが Yokohama Art Spot にアクセスすると, まず図 5 に示す横浜市内の地図と共に, LODAC Museum から表示範囲内に含まれるアート関連施設, ヨコハマ・アート LOD からそれらの施設で開催されるアート関連イベント (展覧会など), PinQA からその周辺で発生した Q&A 情報がマーカーで表示される. この機能は, 地図の表示範囲の緯度および経度の情報を基に各 LOD の SPARQL エンドポイントに SPARQL クエリが送られ, 表示範囲内の情報が返されることで実現している. 図 4 に示すように, LODAC Museum の施設が表示範囲に含まれていれば, その施設が開催場所としてリンクされているヨコハマ・アート LOD のイベント情報は地図に表示される. さらに, 施設情報について詳細を見る場合には, 施設詳細のページに移動する.

\*4 <http://pinqa.com/>

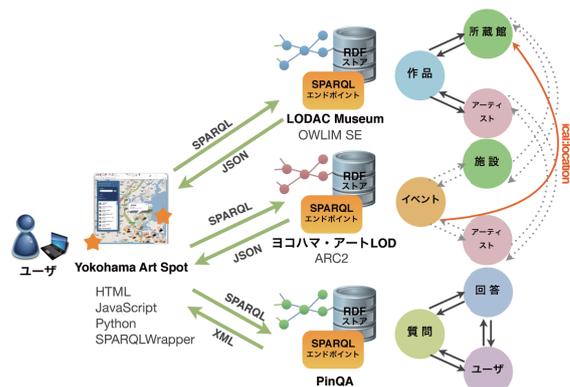


図 4: Yokohama Art Spot の構成

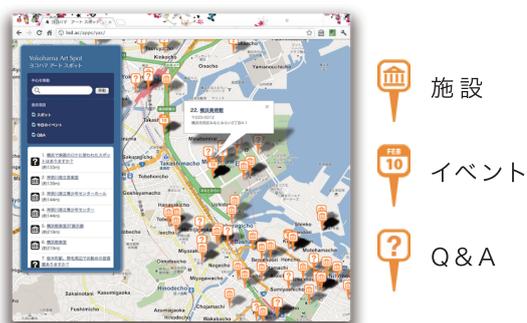


図 5: Yokohama Art Spot マップページ

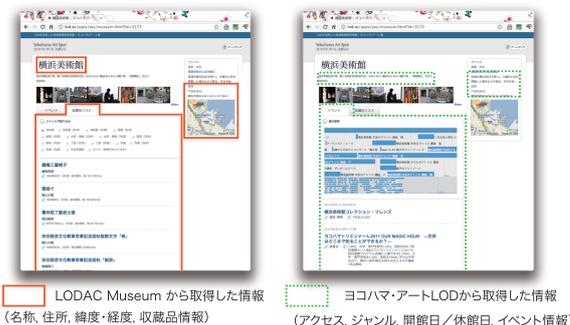


図 6: Yokohama Art Spot 施設ページ

このとき, 図 6 に示すように, 施設名, 住所, 収蔵品リストなど (実線枠内) は LODAC Museum から, アクセス, ジャンル, 開館日, イベント (点線枠内) などは LODAC Museum における当該施設の URI をキーとしてヨコハマ・アート LOD から動的に表示される. 収蔵品リストからは, さらに LODAC Museum のページに移動して詳しいデータを探索することが可能である. これにより, たとえばあるイベントに参加する予定のユーザが開催場所である施設のページを閲覧する際に, 収蔵品リストにも興味を持ち, 思わぬアーティストの作品を発見し, 常設展の鑑賞も行動予定に入れるなど, ユーザの新たな興味を喚起できる可能性がある. このように, 互いにリンクを持つ複数の LOD を併せて活用することで, 次々と蓄積されていくデータを活用する機会の創出が期待される.

### 5.2 ふおとぶらり

ふおとぶらり\*5 は, 株式会社 ATR-Promotions により開発された iPhone/iPod/iPad 用の位置情報に基づく写真共有アプリ

\*5 <http://museum-media.jp/web/ja/products/ふおとぶらり>



図 7: ふおとぶらりによる LODAC データ表示

ケーションである。ふおとぶらりにおいても図 7 に示すように、写真とともに地図上に LODAC Museum から得られたアート関連施設がマッピングされ、その詳細情報を閲覧することが可能である。また、その他にも DBPedia や LinkedGeoData などの汎用的な LOD に対して、LODAC Museum から表示範囲内の施設情報を取得するための SPARQL クエリとほぼ同じクエリを送ることで、それらの LOD が持つランドマークなどの位置情報を取得し、表示することもできる。このように LOD では SPARQL など標準化された技術を利用しているため、API のように仕様がサービスごとに変わず、異なる LOD に対して同じクエリを使える場合があり、新たな LOD の追加などの変化に伴う開発コストが抑えられるという利点がある。また、写真を撮影した場所の周辺情報が簡単に取得することで、撮影時の状況の想起などに役立つと考えられる。

### 5.3 ミュージアムへ行こう！

ミュージアムへ行こう！\*6 は、Twitter ユーザの発言に緯度および経度が含まれている場合、もしくは地名が含まれている場合に、そのユーザの位置に最も近いミュージアムを LODAC Museum から検索し、その概要や経路をユーザに向けてリプライする。さらに Web サイトでは図 8 に示すように、LODAC Museum だけでなく、東日本大震災で被災した図書館、公文書館、博物館、公民館を支援するプロジェクト saveMLAK\*7、Wikipedia から得られる情報も併せて表示することで、より充実した博物館情報を地図上で提供している。このように、LOD と Twitter を連携させることで、ユーザに対して LOD の情報をプッシュすることが可能になった。4 節における Semantic MediaWiki と同様に、Twitter におけるユーザとのインタラクションを用いて、LOD にアノテーションを行うことも期待される。

## 6. まとめ

本稿では、国内の博物館情報の LOD 化を図る LODAC Museum の活動について報告した。LODAC Museum では、これまでの 15 館に加えて、新たに 38 館の博物館の Web サイトからコレクション情報を追加収集し LOD 化した。これらの大規模なデータ収集と LOD 化に伴い、データ処理とメタデータ設計を分離する LOD 構築システムを導入することで、LOD 生成の効率化を図った。また、ユーザが LODAC Museum にデータの追記を要求する場合に備えて、Semantic MediaWiki を導入し、ユーザの定義したメタデータスキーマに基づいて特定のデータにアノテーションすることを可能にした。さら



図 8: ミュージアムへ行こう！ トップページ

に、LODAC Museum だけでなく、ヨコハマ・アート LOD および PinQA を連携活用した横浜のアート情報を提供する「Yokohama Art Spot」や、写真とともに様々な LOD に含まれる位置情報も提示可能な iPhone アプリケーションの「ふおとぶらり」、Twitter と LOD を利用したユーザへのプッシュ型情報提示を行う「ミュージアムへ行こう！」など、LODAC Museum を用いたアプリケーションの開発を行い、LOD の活用により期待される効果について議論した。今後は、ヨーロッパの博物館情報の LOD である Europeana へのリンクの生成や、ユーザによる LOD へのアノテーションおよびデータ追加方法の改良、国内の博物館のコレクション情報のさらなる LOD 化を目指す。

## 参考文献

[Bizer 2009] Bizer, C., Heath, T. and Berners-Lee, T.: Linked Data – The Story So Far, International Journal on Semantic Web and Information Systems (IJSWIS), Vol.5, No.3, pp.1–22, 2009.

[Haslhofer 2011] Haslhofer, B. and Isaac, A.: data.europeana.eu – The Europeana Linked Open Data Pilot, Proceedings of the International Conference on Dublin Core and Metadata Applications (DC2011), pp.21–23, 2011.

[Krötzsch 2006] Krötzsch, M., Vrandečić, D. and Völkel, M.: Semantic MediaWiki, Lecture Notes in Computer Science, 4273/2006, pp.935–942, 2006.

[福田 1997] 福田博同, 五十殿利治: 美術シソーラスデータベース形成の諸問題, 情報管理, Vol.40, No.9, pp.790–809, 1997.

[加藤 2011] 加藤文彦, 嘉村哲郎, 大向一輝, 武田英明: コレクション型ウェブサイトから Linked Data を構築するプラットフォームの開発, SIG-SWO-A1103-01, 第 26 回 セマンティックウェブとオントロジー研究会, 人工知能学会, 2011.

[嘉村 2011] 嘉村哲郎, 加藤文彦, 松村冬子, 上田洋, 高橋徹, 大向一輝, 武田英明: 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, じんもんこん 2011 論文集, 情報処理学会, Vol. 2011, pp.409–416, 2011.

[武田 2011] 武田英明, 嘉村哲郎, 加藤文彦, 大向一輝, 高橋徹, 上田洋: 日本における Linked Data の普及にむけて, 3E3-OS20-9, 人工知能学会第 25 回全国大会論文集, 2011.

[松村 2011] 松村冬子, 小林巖生, 嘉村哲郎, 加藤文彦, 高橋徹, 上田洋, 大向一輝, 武田英明: Linked Open Data による博物館情報および地域情報の連携活用, じんもんこん 2011 論文集, 情報処理学会, Vol. 2011, pp.403–408, 2011.

\*6 <http://twitter.com/go2museum>  
 \*7 <http://savemlak.jp/>