

[B12] Wikimedia Commons を利用したデジタルアーカイブ公開の試み: 東京藝術大学 音楽学部大学史史料室所蔵史料を例に

○嘉村哲郎¹⁾

¹⁾ 東京藝術大学 芸術情報センター, 〒110-8714 東京都台東区上野公園 12-8

E-mail: kamura@noc.geidai.ac.jp

An attempt to create digital archives using GLAM-WIKI methodology: A case of the GEIDAI Archives

KAMURA Tetsuro¹⁾

¹⁾ Art Media Center, Tokyo University of the Arts, 12-8 Uenopark, Taito-ku, Tokyo, 110-8714

【発表概要】

中小規模のデジタルアーカイブの課題の一つに、データ保存と管理に必要な情報システムや機器の継続性があげられる。そこで、情報システムやメンテナンス費用、運用にかかる属人的要素の課題を解決する案として、オープンライセンスでデータを Web 公開・共有を可能とする Wikimedia Commons と Wikidata の利用を提案する。本稿では、東京藝術大学音楽学部大学史史料室の資料を例に、これら Wiki 群を使用したデータ公開と活用『GLAM WIKI』について報告する。

1. はじめに

東京藝術大学音楽学部大学史史料室では、総合芸術アーカイブセンター（2011 年 5 月～2016 年 3 月）の活動以来、所蔵資料のデジタル化を進めている。デジタル化したデータの保存と Web 公開は、標準耐用年数を大幅に超過した当時のアーカイブセンターが導入した機器を継続利用している。本学史料室のように、小規模なデジタルアーカイブの現場では慢性的な資金不足に悩まされており、予算や人員削減の理由から情報機器の運用・管理、データのマイグレーションが困難な状況が散見される。そこで、著者らは情報システムのメンテナンスやデータ保存のためのストレージ、Web 公開に係る諸費用の問題を解決するために、Wikimedia 財団が無償で提供している Commons (<https://commons.wikimedia.org/>)と Wikidata (<https://www.wikidata.org/>)を用いたデータ公開の活動 GLAM WIKI に着目した。本稿では、東京藝術大学音楽学部大学史史料室の資料（吉本光蔵撮影日露戦争写真）を対象に、実際にこれらの Wiki を用いた画像公開や構造化データの作成方法を紹介し、本取組で明らかになった課題点等を報告する。

2. GLAM WIKI とは

GLAM WIKI は、芸術や文化に関連する情報を扱う組織 Galleries, Libraries, Archives, Museums が保有するコレクション情報やメディアデータを、Wikimedia 財団が提供する Wiki システムを通じて公開・共有する活動で

ある。その発端は、WWW の発明者ティム・バーナーズ＝リー卿が 2009 年に呼びかけた“Raw Data Now” [1]を機に様々なオープンデータ活動と共に起こったと考えられ、2010 年には GLAM WIKI の名を冠したイベントが行われている[2]。また、関連する活動には Open GLAM がある。こちらも GLAM WIKI と同様に、GLAM が保有する情報を二次利用可能なデータで公開・共有を促進する活動であり、2011 年にはワークショップが開催され、現在も世界中で活動が続いている。

ここ数年は、特に GLAM WIKI の活動が注目を集めている[3]。その背景には幾つかの要因が考えられるが、芸術や文化に関連するデータをオープンデータで公開するためのライセンス、Creative Commons の成熟や Rights Statements 等新たな指針の考案と普及、Wiki システムでデータ公開するためのツールの開発が進み Wiki 利用の利便性が向上したこと、データの再利用性を考慮した構造化データが API 経由で利用できるようになったこと、米国のメトロポリタン美術館に見るように、著名な美術館・博物館が Wiki を基盤としたデータ公開と活用の取組が相次いでいることが考えられる。さらに、Wiki システムは無償で利用できるため、情報システムへの投資が厳しくてもデジタルアーカイブを公開したい小規模の GLAM 組織は、ストレージ機器やデータ公開用のシステム費を負担することなく利用できるため、このような組織が Wiki を活用することは、費用面で大変効果的である[4]。

3. GLAM WIKI とデータ公開

Wikimedia Commons (以後 Commons という) と Wikidata を用いてデジタルアーカイブを公開する方法, すなわち GLAM WIKI でデータ公開をするにあたり, Commons と Wikidata の特性を理解しておく必要がある。前者は, Wikipedia と同様に Wiki システムをベースにした画像, 音声, 動画等のメディアデータの保存と公開に特化したストレージ的役割の機能を持つ。公開データに記述できるメタデータは, Wiki の記法に準拠したテンプレートが用意されており, データの種類に応じてメタデータが設定されている。一方の Wikidata は, Wikimedia 財団が提供する Wiki 群(Wikipedia や Wikibooks 等)と呼ばれる Wiki システムの情報を構造化データで共有・公開することを目的とした, あらゆる情報がオンラインで利用できる知識データベースである。Wikidata の特徴は, すべてのデータが Linked Open Data で提供されており, 専用の API (SPARQL Endpoint)を用いてデータ利用できる点にある。

2 つの Wiki システムを用いたデジタルアーカイブの公開手順は, 大まかに以下の通りである。本稿では既に 1, 2 があつた事を前提に 3~5 の取組みを紹介する。

1. 公開画像データと目録リストを用意する。
2. Wikimedia の利用者アカウントを用意する。
3. Wikimedia Commons への画像公開。
4. Wikidata 用の構造化データ作成。
5. Wikimedia Commons のギャラリー機能, または Wikidata の API (SPARQL)を用いた組織 Web でコンテンツを公開する。

3.1 Wikimedia Commons への画像公開

Commons への画像公開は, ファイル 1 点毎またはツールを使用した複数画像を一括アッ

プロードする方法がある。本稿では, 51 件の画像を扱うため。公式サイトが推奨するツール Pattypan (<https://github.com/yarl/pattypan>)を用いた。Pattypan を起動後すると, 「Generate Spreadsheet」と「Validate & Upload」の選択が表示される。はじめに, Commons 用のデータリストを作成するため, 前者を選択して公開画像のディレクトリを指定すると, メタデータを設定するテンプレート作成画面になる。今回は, 写真データを扱うため, Commons の Photograph テンプレートを用いた。その後, 必要なメタデータを選択し, 入力欄には各画像が共通する内容を入力する (図 1)。この時, 組織名や関連人物, 作者名等の情報が Wikipedia にある場合, その情報を記載する。例えば, Wikipedia 上のリソース, 東京藝術大学を Commons に関連付ける場合は, `{{Institution:Tokyo University of the Arts}}` の様に Wiki の記法で入力する。必要項目を入力後, 画面を進めると最終的に EXCEL 形式のリストデータが生成される。リストデータには, 画像毎のタイトルや解説文等の情報を追記する。なお, リストデータの 2 シート目には Wiki テンプレート情報が格納されており, 必要に応じて項目を追加・削除することができる。リストデータを編集・保存後, 改めて Pattypan を起動し「Validate & Upload」からアップロード前の確認を行う。アップロード前の確認は, ツール上の確認と Commons 上で実際の表示例を確認できる。内容を確認後, Wikimedia のアカウントとパスワードでログインし, すべての画像とメタデータをアップロードする。Commons にアップロードしたデータは, 個々の URL を有し, 各画像を閲覧できる他, 多言語表記に対応した専用ビューアによるスライドショーが利用できる (図 2)。

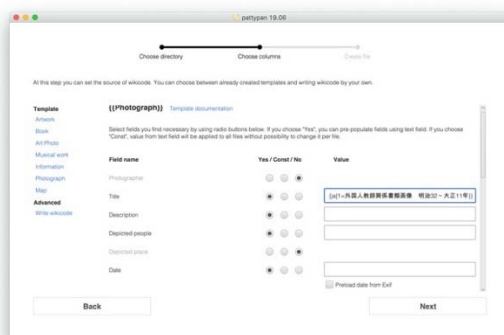


図 1. Commons のテンプレート作成画面



図 2. Commons に公開した画像のスライドショー

4. 課題点

次に、Commons と Wikidata の公開を通して、それぞれの課題点を述べる。

① データ内容の編集について

Commons は、誰でも自由に編集できる Wiki という特性上、大量の情報を一括更新したい場合は、Bot と呼ばれる専用のプログラムと特殊なアカウントが必要になる。そのため、複数の情報を更新するためには特殊な環境と専門性が求められる (データ 1 点毎の更新は手動で可能)。一方、Wikidata は、内容を更新する際も Open Refine を用いることで、既存の情報を上書きすることができる。この点から、将来的に情報の修正や更新作業が発生することを考慮すると、Common に記載する情報は最小限に留めておき、詳しい内容は Wikidata で管理した方が良いと感じた。ただし、Wikidata が標準で用意している「解説」プロパティは、250 文字の制限があるため、リソースの解説に長文が必要な場合は、別途プロパティを追加するなどの対応が必要になる。

② データ利用について

Common は、メディアストレージとしての機能を有し、専用の API で公開データを利用できる。しかし、データ検索範囲は Commons カテゴリやテンプレートに依存するため、限られた範囲となる。一方の Wikidata は、LOD の仕組みを備え、SPARQL Endpoint が利用できるため、データのリンク構造を辿る検索や、検索結果をそのままデータとして二次利用する等、柔軟なデータ活用が可能である。さらに、Wikidata では SPARQL クエリの作成を支援する Query Service (<https://query.wikidata.org/>) の提供と、検索結果を組織の Web サイトで公開できる機能を備えている。データ活用の点から、Commons より Wikidata を中心に内容を充実することが好ましい。しかし、Wikidata にも短所はある。SPARQL で Commons のデータを扱うためには、Commons と Wikidata の双方に情報が必要になる (2020 年 2 月末時点)。例えば、Commons の日露戦争カテゴリには約 800 点のデータがあるが、その多くが Wikidata にリソースとしての登録がないため扱えない。

5. おわりに

最後に Wikidata の活用例と可能性を述べる。図 6 は Wikidata の SPARQL を利用して「日露戦争」に属するリソースの画像を抽出し、大学史料室の Web サイトで日露戦争関連写真として公開したものである。結果、従来は吉本光蔵撮影日露戦争写真のみの情報量であったが、Wikidata の利用によりこれまで知られていなかった日露戦争関連の情報を提供できるようになった。Wikidata を用いたデータ活用の利点は、従来のデータベースでは困難であった Web に公開されたデータから API 経由で新たな知見を獲得できるようになった点である。今回は、単純な検索クエリ例を示したが、撮影場所や資料館情報から所蔵国や資料数の算出、地図に位置情報を表示するなど、新たな資料データの活用可能性が伺えた。

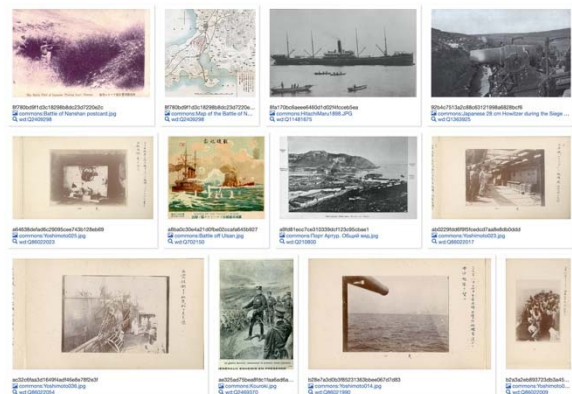


図 6. SPARQL による「日露戦争」関連写真の抽出

参考文献

- [1] Tim Berners-Lee, The next web, 2009.
https://www.ted.com/talks/tim_berners_lee_the_next_web
- [2] GLAM WIKI, 2010.
https://wikimedia.org.uk/wiki/GLAM-WIKI_2010
- [3] Valeonti, F., Terras, M., & Hudson-Smith, A. (2019). How open is OpenGLAM? Identifying barriers to commercial and non-commercial reuse of digitised art images. *Journal of Documentation*, 76(1), 1–26.
- [4] Elizabeth Joan Kelly, ‘Assessing Impact of Medium-Sized Institution Digital Cultural Heritage on Wikimedia Projects’, *Journal of Contemporary Archival Studies*, 6.1 (2019), 25.



この記事の著作権は著者に属します。この記事は Creative Commons 4.0 に基づきライセンスされます (<http://creativecommons.org/licenses/by/4.0/>)。出典を表示することを主な条件とし、複製、改変はもちろん、営利目的での二次利用も許可されています。