

公開データとリポジトリソフトウェアを利用した簡易的な標本データベースの構築

著者	吉岡 翼
雑誌名	富山市科学博物館研究報告
号	45
ページ	19
発行年	2021-07-01
URL	http://repo.tsm.toyama.toyama.jp/?action=repository_uri&item_id=2034

資 料

公開データとリポジトリソフトウェアを利用した 簡易的な標本データベースの構築*

吉岡 翼

富山市科学博物館

939-8084 富山市西中野町一丁目8-31

Development of a Collection Retrieval System Using the Open Data and Repository Software

Tasuku Yoshioka

Toyama Science Museum

1-8-31 Nishinakano-machi, Toyama 939-8084, Japan

1. はじめに

標本データのウェブ公開ツールは、資料管理システムの一機能や専用のウェブアプリケーションが存在するほか、汎用的なウェブコンテンツ管理システム（以下CMS）においてもプラグインなどを利用して標本データの公開機能を持たせることが可能である。しかしながら、こうしたシステムを構築し安定的に運用するには様々な障害も存在する。初期の構築段階では予算の確保やシステム設計、データ入力が必要なハードルとなりえ、安定的な運用には技術の共有や継承、システム移行時の持続性保持といった課題もある。

富山市科学博物館（以下当館）では標本情報の電子化は早い時期から行われているが（太田, 1987, 1991）、分野によって管理手法や電子化の進捗が異なり、標本データへの電子的なアクセス手段は独自に提供できていなかった。一方で国立科学博物館を介したS-Net/GBIFへの標本データ提供のため、収蔵されている半数以上の自然史標本約25万点分の情報が既に正規化され、ウェブ上で流通している。

このような背景の中、当館では機関内部で生み出された学術情報の電子的な集積・公開を目的として、CMSベースの機関リポジトリを2018年に導入した。リポジトリの収録対象は主として発行物などの文献資料であるが、標本データもシームレスに参照できると便利である。そこで、同一のシステムをベースに標本データベースを構



図1. 標本データベースのトップページ

築し、S-Net/GBIFの当館分データを収録した（図1, 2）。また、持続的なアクセスを確保するための手段や、QRコードを用いて標本データに容易にアクセスする仕組みも並行して整えた。本稿ではこれらのシステム構築について報告する。

2. 標本データベース

当館では機関リポジトリとして国立情報学研究所で開発されたWEKO2を独自サーバで運用している。WEKO2はBSDライセンスで公開されており、標本データベース用に一部改変して利用した。博物館資料のような非文献資料をリポジトリ公開する取り組みは以前からあり（堀井ほか, 2012; 福原, 2012）、標本データベースという形で公開している例としても島根大学総合博物館の「島根大学標本資料データベース」があげられる。リポジトリでは電子コンテンツがそれを記述するメタデータとともに格納されるが、純粋な標本データベースとしてならば、電子コンテンツでなく物理的実体に対して記述したデータを格納するという構造が求められるべきである。物理的実体を上位階層にするなどの操作を行えば合理的な実装もできるが、改修が大掛かりとなるのでそこまでは求めず、柔軟なデータ項目の設定や、画像等の電子ファイルの関連付け、データの一括登録などが容易に行えるシステムとして、リポジトリシステムを標本データベースとして利用することとした。そのため、リポジトリシス

* 富山市科学博物館研究業績第584号

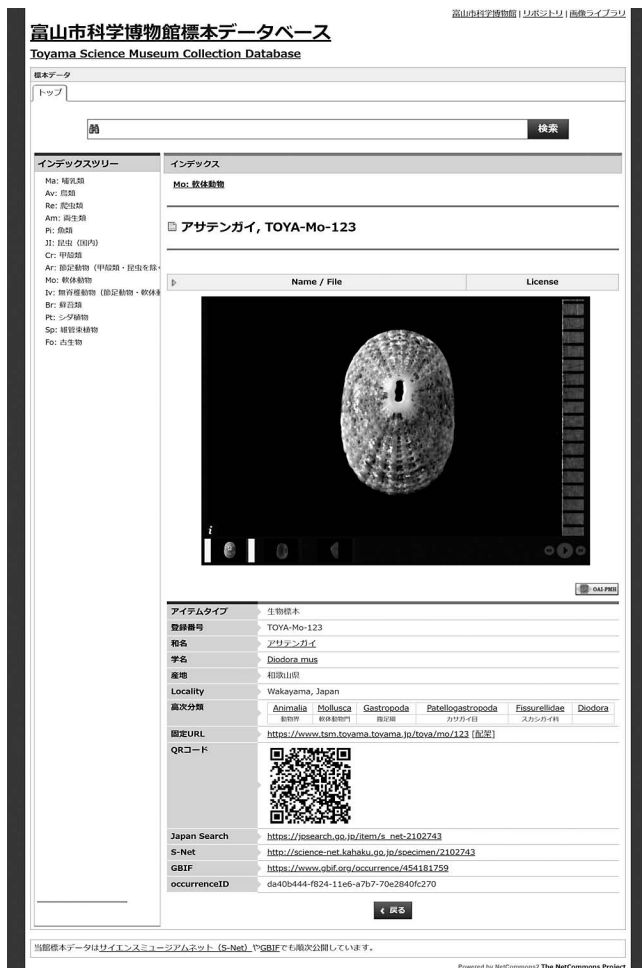


図2. 個別の標本情報画面。

テムを用いているが、基本的に本体ファイルを持たない形で運用している。当館の機関リポジトリ本体とは性質が異なるため運用上のシステムは切り離しているが、既存システムを活用しているので構築のハードルは低く抑えられている。

標本に関するデータはS-Net/GBIF用に正規化されたデータおよび、GBIFで既に公開されているデータから抽出して利用した。GBIFの標本データはダーウィンコアに準拠しており多数の項目を持つが、このうち表示させたい最小限の項目のみ標本データベースに収録した。分類群や地名など階層構造をもつ項目は区切り文字を含む文字列として結合して1つにまとめた。化石標本では生物標本にはない層序や地質年代の情報を持つように、対象によって記述する項目が異なるので、アイテムタイプを利用して種別ごとに項目を設定した。今回はデータ交換を前提としていないが、可能な範囲でダブリンコアの基本要素にマッピングし、OAI-PMHでのハーベストもできる状態にしてある。リポジトリへのデータ登録にはSWORDというプロトコルが用意されているが、当該環境では接



図3. 標本ケースに表示したQRコード。



図4. キャビネットに表示したQRコード。

続に失敗したため、事前に作成したXMLファイルを用いてブラウザ経由で登録した。ブラウザ経由の登録には制約もあったが、1人で可能な程度の労力ですんだ。

S-Net/GBIFに存在しないデータとして、標本画像などの電子ファイルと標本を参照している文献の情報を入力できるようにした(図2)。画像や文献の情報追記はデータ作成と平行するので、編集画面から個別に少数ずつ行った。参考文献は検索対象のテキストにDOIなど文献の識別子を含め、同一文献中で参照されている標本を検索で一覧できるようにした。また、機関リポジトリ本体においても文献内で扱っている標本データを検索できるリンクボタンを設置し、相互に参照できるようにした。

将来システムを移行する際は現存データを移植する方法もあるが、S-Net/GBIFのデータが最新の状態で更新されていくのであれば、S-Net/GBIFのデータを公開用マスターとして再び取り込むことも想定できる。画像や文献の情報のように後付けのデータは個別に管理する必要があるが、大半のデータがシステムに依存しない形で存在するので、技術の共有や継承が固有の課題とはなり

にくい。

3. 持続可能で視認性の高いURLとグループ化

自然史標本はそれを特定・管理するために登録作業を行って番号を与え、対外的にも標本を参照する際の識別子として用いる。多くの場合、機関コード、コレクションコード、登録番号の組み合わせで表記される。CMSで生成される個別標本のURLも識別子として用いえるが、普段使っている標本番号とは無関係なパラメータの羅列となり、標本情報へのアクセスには検索フォームを使うなどの手間が必要となる。また、システムを移行した際にはURLが変わってしまい、資料情報への持続可能なアクセスが困難となるリスクがある。そこで、コレクションコードと資料番号からなる視認性の高いURLを固定URLとして、標本データが表示されるページにリダイレクトできるように設定した。固定URLは以下のようなものである。

<https://www.tsm.toyama.toyama.jp/toya/mo/123>

上記例のmoはコレクションコード、123は資料番号である（当該資料は軟体動物コレクションに含まれる和歌山県産のアサテンガイ；図2）。論文等の記事中で資料を特定する際は機関コード（当館はTOYA）と併せてTOYA-Mo-123のように示すので、URLと記事中の識別子を容易に対応させられる。リダイレクト用のデータセットは標本データベースで用いているデータテーブルから作成した。

また、アクセスには制限を設けているが、標本を配架するキャビネット等、複数の標本等が所属するグループにも識別子を与え、グループ内標本の参照や、逆に標本から所属するグループを参照できるようにした。例えば、先のアサテンガイの配架情報は以下のURLから参照可能となっている。

<https://www.tsm.toyama.toyama.jp/toya/mo/123.arr>

4. QRコードの付加

古生物および貝類資料から順次試みている段階だが、ウェブ上の標本情報へより容易にアクセスするため、標本の容器やキャビネット等の見えやすい箇所に、固定URLを格納したQRコードをシール等で表示した（図3, 4）。QRコードが読み取れる携帯端末は普及しているので、収蔵庫や借用先等で標本情報へ容易にアクセスすることが可能である。また、固定URLがコレクションコードと資料番号の組み合わせになっているので、アプリケーション側で対応すればリストを作成するなどウェブ表示以外の処理にも利用しやすく、将来の管理の効率化が期待できる。また、データベースに未収録の標本に対しても先

行してラベリングが可能である。QRコードはウェブ上の標本情報ページにも表示させており、標本を前にした操作を模擬的に行うこともできる。

近年はQRコードなどの画像による自動認識に代わって電子タグ（RFID）も普及しているため、QRコードとともに採用を検討した。電子タグは情報密度がより高く、標本ラベルや台帳上の情報をそのまま格納することもできる。非接触で情報の追加・変更ができ、低価格化してきていることから、図書館資料の管理には普及しているし、博物館資料の管理にも有望と思われる（高木, 2004）。しかし、自然史資料は大きさや形状・材質が様々で、直接タグを貼り付けることができない場合が多い。既存の文字による標識や標本ラベルについても収納容器への同梱や容器に貼るといったことが一般的である。腊葉標本のように台紙等に固定するものを除くと、標本と電子タグは別々になりやすく、資料管理を確実に行うためには結局標本の存在を目視して確認せざるを得ない。識別子のみを格納したQRコードでもネットワーク環境があれば標本情報にアクセス可能であり、電子タグを採用してもメリットは十分に享受できない。また、QRコードを印刷したシールに比べるとコストもかかるため、今回は電子タグの採用は見送った。

5. おわりに

まだ試行的な部分が多く、システムの完成度も高いとは言えないが、すでにあるデータや汎用されている仕組みの組み合わせにより、将来のシステム移行も織り込んで構築することができた。今後収蔵庫の整備を進めながら、より利用しやすい仕組みに発展させていきたい。

6. 引用文献

- 福原 正, 2012. CURATORによる萩庭標本データベースの公開について. *CURATOR Letter*, (7): 1.
- 堀井 洋・堀井美里・林 正治・塩瀬隆之・高田良宏・古畑 徹, 2012. 分野・組織横断的な非文献資料リポジトリの実現を目指して. *情報知識学会誌*, 22(2): 91-96.
- 太田道人, 1987. 富山市科学文化センター収蔵資料目録 1号 進野久五郎植物コレクション. 富山市科学文化センター.
- 太田道人, 1991. 植物標本データベース構造の一提案. 富山市科学文化センター研究報告, (14): 79-86.
- 高木博彦, 2004. 非接触ICタグ（RFID）を用いた博物館資料の管理—予察—. 千葉県立現代産業科学館研究報告, (10): 89-39.