

# SAT 大正蔵画像 DB をめぐるコラボレーションの可能性

永崎研宣<sup>†1</sup> 津田徹英<sup>†2</sup> 下田正弘<sup>†3</sup>

**概要:** 本稿は、2015 年度に Web コラボレーションによってタグ付けされ 2016 年春に IIF に準拠して公開された SAT 大正蔵画像 DB の開発と公開に関する報告であり、一連の活動と今後の課題におけるコラボレーションの可能性について検討している。

**キーワード:** IIF, 仏尊画像, 高精細画像, Mirador

## Collaboration for SAT Taishōzō Image DB

Kiyonori Nagasaki<sup>†1</sup> Tetsuei Tsuda<sup>†2</sup>  
Msahiro Shimoda<sup>†3</sup>

**Abstract:** This manuscript describes SAT Taishōzō Image DB which was tagged during 2015 and released compliant with IIF in the spring of 2016. It mentions possibilities of collaboration in the activities and future of the project.

**Keywords:** IIF, Buddhist images, hi-resolution images, Mirador

### 1. はじめに

本稿では、SAT 大正蔵画像 DB (以下、SAT 画像 DB) の開発と公開について報告し、SAT 画像 DB をめぐるコラボレーションの今後の可能性について検討したい。

### 2. 大正新脩大藏經画像編

『大正新脩大藏經』は、100 巻から成る仏典の集成であり、テキストを主とする 85 巻分に関しては、2007 年には SAT 大藏經テキストデータベース研究会[a] (代表: 下田正弘) (以下、SAT プロジェクト) によってフルテキストデータベース (以下、SAT テキスト DB) [b]として公開され、様々に活用されてきている[1]。それ以外の 15 巻分、とりわけ、画像編・画像部等と呼ばれる 12 巻分に関しては、テキストのみならず、仏尊や三昧耶形・曼荼羅等の画像を多く含んでいる貴重な資料である。したがって、テキスト検索できるよりはむしろ、ここに含まれる画像を何らかの形で検索できるようにすることが希求されていた。そこで、SAT プロジェクトでは、この 12 巻分を高精細画像としてデジタル撮影し、画像の属性に関するタグを付与して検索できるようにすることで利便性を高めることを目指した。

### 3. Web コラボレーションシステム

タグの付与は、全国に散らばる日本美術史の研究者によって行うこととなった。これには、画像上でタグを協働でつけていくシステムが有益であり、Web コラボレーションシステムを開発することとなった。当初は、Annotorious[c]

という Web アノテーションツールを基本に、画像の一部にタグをつけるという仕組みを構築した。タグの内容については、髪部、面相、臂数、持物、台座、眷属等、仏尊の部分ごとに属性のリストを作成し、作業者は、名称以外に関しては、Web フォームでクリックして選択するだけで入力できる仕組みとした(図 1)。サーバ側は、PostgreSQL 上に、一つの画像に対して一つのレコードとして、一レコード内に複数のタグを保存していく仕組みとした。作業員から見ると、タグを付与したい領域を画像上で選択すると入力フォームが表示されるので、そこに仏尊名を入力し、その後、さらに属性リストのフォームを開いて該当する属性をクリックして選択し、送信ボタンをクリックすると、一つのタグ付け作業が完了となる。共同作業を適切に統制するため、これまで SAT プロジェクトが開発・利用してきた他のコラボレーションシステムと同様に、作業員権限以外に管理者権限を用意し、管理者のみが、他の作業員・管理者のタグを変更することや作業員毎の作業進捗状況のチェック、さらにはアカウント情報の管理もできるように管理者用のインターフェイスも用意した。また、作業に際して、タグの用法などの問合せにかかる手間を減らすため、当初より検索機能を提供し、付与したタグ名で検索できるようにしつつ、検索結果は IIF[d] Image API によって座標情報を用いて画像の該当部分を切り出してリスト表示できるようにした。これにより、作業員は、タグの用法について画像を検索して確認しながら作業できるようになり、作業効率を高めることができた。最終的に、この作業には 43 名の研究者

a) <http://21dzk.l.u-tokyo.ac.jp/SAT/>

b) <http://21dzk.l.u-tokyo.ac.jp/SAT/satdb2015.php>

c) <http://annotorious.github.io/>

d) <http://iif.io/>

が参加した。



図1 タグ付け用フォームの画面

#### 4. 公開に向けて：IIIFの採用

タグ付け作業は2015年度に実施され、最終的には画像編第1巻・第2巻に登場する4300件の仏尊・三昧耶形等にタグを付与することができた。この公開にあたっては、IIIF (International Image Interoperability Framework)のImage API及びPresentation APIが2.0となり対応する各種ソフトウェアもこれに対応したことで規格として安定し、海外各地のデジタルリポジトリにおいて採用が進みつつあり、さらに、SAT 画像DBの公開に適したフリーのWebビューワも公開されていたことから、公開に際してIIIFを採用することとした。

IIIFは、いわゆる文化資料デジタルアーカイブにおける画像配信の手法を共通化して低コスト化しつつ利便性を大幅に向上させることを目指したものである。元々は、世界に散らばる中世写本のデジタル画像を統合したコレクションを作ることを課題としたプロジェクト[e]が発展してきたものだが、現在では、欧米・中東などの研究図書館におけるデジタルアーカイブを中心に、一気に普及しつつある。この規格では、「サイロを越える」等のスローガンの下、各地の「デジタルアーカイブ」サイトにおける写本・版本等の画像資料を効率的にまとめてコレクションとして提示したり共同で翻刻やアノテーションを付与したりすることができるようになってきている。また、IIIFに準拠すればWebビューワを共通化できるため、これをターゲットとした利便性の高いフリーソフトのWebビューワがいくつも開発公

開されるようになってきている。とりわけ、Mirador[f]とUniversal Viewer[g]は、多機能なIIIF対応ビューワとして広く用いられている。いずれも、高精細画像に関してはOpenSeadragon[h]を利用しており、他にも様々なツールやライブラリを組み合わせられて作成されている。そして、前者は、今のところ対象は高精細画像のみだが、一つのウィンドウに複数のスロットを作り、それぞれに拡大縮小表示とアノテーション表示が可能となっている。IIIFの当初の目的の一つであった複数の写本・版本を子細に見比べたいというニーズはよく合致していると言える。現在リリース候補が出ているバージョン2.1ではアノテーションを付加する機能や画像の回転や明度・彩度などの調整もブラウザ上でできるようになっている。サーバ側がIIIFに準拠していれば、どこのサーバの画像でも同じように表示して操作できることから、Miradorの複数画像表示機能はIIIFの特性をよく活かしたビューワであると言えるだろう。また、Universal Viewerは、3Dや音声・動画など、様々なタイプのメディアを表示する機能を備えている。XMLテキストの検索機能も用意されており、こちらもMiradorとは異なる方向での発展が期待されることである。

このような状況において、SAT 画像DBでは、Miradorの複数画像表示機能の利便性に着目し、これを活用してSAT研究会の考える利便性の高い表示機能を提供することとした。これにより、IIIF対応としてImage APIとPresentation APIに準拠する形で画像とアノテーションを公開することになった。さらに、SAT 画像DBでは、CC BY-SAのライセンスで公開したことで、他のビューワや、あるいはまったく別の仕組みでの共有・公開など、様々な選択肢をも利用者に提供することとなった[i]。

#### 5. IIIFの実装と付加的な機能

IIIFを書籍のデジタル画像に対して実装するにあたっては、まず、画像そのものを様々な操作するIIIF Image APIを利用するための画像配信サーバソフトをWebサーバ上の、高精細画像が読み出せるところに用意する必要がある。そして、一つもしくは複数の画像をまとめて操作するIIIF Presentation APIを提供するためにJSONデータの作成が必要になる。

画像配信サーバソフトに関しては、SAT 画像DBでは、サーバ環境の制約からLoris IIIF Image server[j]とIIP Image server[k]の二つを検討し、SAT 画像DBで公開予定のJpeg画像が最大で80MBに達していることから、動作そのものが比較的速く、Pyramid TIFFを扱うことで配信時の画像分

e) <http://www.slideshare.net/Tom-Cramer/iiif-international-image-interoperability-framework-df12012>

f) <http://projectmirador.org/>

g) <http://universalviewer.io/>

h) <https://openseadragon.github.io/>

i) 本稿執筆時点で、外部でこのデータを利用しているサイトとしては神崎

正英氏による以下のものが挙げられる。

<http://www.kanzaki.com/works/2016/pub/image-annotator?u=http://dzkings.l.u-tokyo.ac.jp/iiif/zuzoubu/01/manifest.json#p9>

j) <https://github.com/loris-imageserver>

k) <http://iipimage.sourceforge.net/>

割の負荷が少なく、memcachedを利用して配信時の分割画像をキャッシングすることでサーバの負荷を減らすことができる IIP Image server を採用することとした。なお、IIP Image server では Jpeg2000 も利用することは可能だが、これに使用できる Jpeg2000 用ライブラリがフリーソフトではないので、SAT 画像 DB では採用を見合わせた。

Presentation API に関しては、SAT 画像 DB で用いている機能は、基本的な図書のメタデータの提供、一つの資料の中での画像の並び順の指定、各画像へのアノテーション、である。ここでは、画像のファイル情報と PostgreSQL に蓄積されたタグデータと、別途用意した目次情報ファイルを読み込んで Presentation API に準拠した JSON データを作成するスクリプトを組んで、静的な JSON データを用意している。アノテーションには HTML タグを用いることができるため、タグやその他の情報は HTML で書き込み(図2)、各タグをクリックしたら画像を検索する、という機能を jQuery 経由で提供している。



図2 タグの表示

画像検索とその結果表示に関しては、IIIF ではなく外部機能として実装を行った。画像検索は、PostgreSQL のテーブルに格納されたタグのテキストを検索することによって実現しており、結果表示には、各タグが付与された画像ファイル名と画像上の座標情報を取得して Image API で同サイズに画像切り出し表示をするようになっている(図3)。この検索結果表示画面は、GET Method により外部から取得できるようになっており、SAT テキスト DB や、国文研データセット簡易 Web 閲覧の『仏鬼軍』においても、この検索結果表示画面が表示されるようになっている。



図3 検索結果の切り出しリスト表示例

検索結果表示画面では、各切り出し画像にチェックボックスが付与されており、チェックをつけた複数の画像を、Mirador の画面分割機能を用いて一度に並べて表示し、それぞれに、対象となった切り出し画像が各画面中央に位置しつつ、ズームしたりタグ表示・タグ検索したりできるような機能も提供している(図4)。



図4 Mirador の画面分割機能による表示

タグ検索機能では、タグ付与の際に用いた語彙をそのまま援用する形で検索フォームを用意し、チェックボックスをクリックすれば検索語が検索語欄に入力され、そのタグで検索できるという機能も提供している。これらの機能は、単に研究上の様々な分析に利用できるだけでなく、仏尊の名称がわからない時に部分的な部位で検索して確認することができるという点でも利用者から評価されている。さらに、仏尊の各部位についての知識が十分でない人や、日本語の入力があまり得意でない人にとっても、仏尊の検索が容易にできるようになり、日本語圏・漢字文化圏以外の人々にとっても好評な機能である。さらに、SAT 大蔵経テキストデータベース[1]において提供している CHISE[m]の漢字オントロジーのデータを用いた漢字の曖昧検索機能と、

l) <http://21dzk.l.u-tokyo.ac.jp/SAT/satdb2015.php>

m) <http://www.chise.org/>

Digital Dictionary of Buddhism[n]を用いた英単語⇒仏教用語変換機能をここでも提供している。これらの機能は、日本語圏以外の利用者から好評を得ている。

## 6. 終わりに：コラボレーションの可能性

最後に、コラボレーションという観点から、今後の課題についてまとめてみたい。まず、タグ付け作業としてのコラボレーションの発展的継続である。曼荼羅画像等、大きな図像に対する細部へのタグ付けが 2015 年のタグ付けシステムでは困難だったため、OpenSeadragon で画像を拡大縮小しつつタグを付与できる仕組みを OpenSeadragon Selection プラグインを用いて新たに開発した。まだ第一巻と第二巻しか本格的なタグ付けを行っていないため、今後は、このシステムを用いてタグ付け作業を進めていく予定である。

また、この図像編には、図像だけでなく、図像を解説するテキストも多く含まれている。このテキスト部分の翻刻も課題となっている。テキスト翻刻に関しては、筆者らは 2005 年から Web コラボレーションで取り組んできており、近年はコラボレーションによる取組みが世界的に広まりつつあるところだが、SAT 図像 DB に関しては、基本的には OCR を中心として進めて行く事を目指しており、OCR 共同校正システムを開発中である。ただし、対象資料は多くが活字であるものの、返り点がついているテキストが多く、一部には写本の画像も含まれており、やや時間をかけてじっくり取り組んでいく必要がある。これに関連して、テキストデータと IIIF のアノテーションとをどのようにリンクしていくかということも課題として取り組んでいるところである[2]。

コンテンツに関するコラボレーションとしては、IIIF に対応して API を公開することで門戸を開いており、さらに、上述のように GET Method で画像検索結果を取得できるようにしている。こうした形でのコンテンツの提供は、プロジェクトの独立性を保ちながら相互連携によりコンテンツの利活用性を高めることにつながっており、今後もこの方向でのコラボレーションを進めていきたい。

IIIF は、規格としても対応ソフトウェアとしてもまだ発展の余地が少なくない。筆者らは、IIIF コンソーシアムへの加盟等、そういった観点からのコラボレーションにも取り組んでおり、IIIF のさらなる発展を視野に入れつつ、適宜、状況に適した技術を採用し、より良いサービスを展開して、関連する研究の発展につなげていきたい。

**謝辞** 本稿は、タグの入力に取り組んでくださった研究者の方々のご協力の賜物であることをここに記して感謝したい。本稿の一部は、国立情報学研究所公募型共同研究「文

化資料デジタルアーカイブの研究活用を志向するフレームワークの研究」の助成、および JSPS 科研費(JP15HP8001, JP15H05725)を受けて遂行されたものである。

## 参考文献

- [1] 永崎研宣. SAT 大蔵経テキストデータベース 人文学におけるオープンデータの活用に向けて. 情報管理 Vol. 58 No. 6 pp. 422-437. 2015.
- [2] Kiyonori Nagasaki, Tetsuei Tsuda, Charles Muller, Masahiro Shimoda. Tagging on Buddhist Images via IIIF and TEI encoding. TEI Conference and Members' Meeting 2016 Book of Abstracts. Vienna (Austria), (2016), pp. 141-143.

n) <http://buddhism-dict.net/ddb/>