

# ウェブアノテーションを用いた仮想コレクション定義の試み -二つの「幸田文庫」を事例に-

林正治（一橋大学 情報基盤センター） 夏目琢史（一橋大学 附属図書館）

松田訓典 山本和明（国文学研究資料館 古典籍共同研究事業センター）

赤木完爾（慶應義塾大学 法学部）

本研究ではウェブアノテーションによる資料画像の分類・整理手法について述べる。とくに、一橋大学附属図書館および慶應義塾図書館が所蔵する「幸田文庫」を対象にしたウェブアノテーションによる仮想コレクション定義手法の提案、実装、その実現に向けた課題を明らかにする。

## Virtual Koda Collection: an attempt to define a Virtual Collection using Web Annotation Model

Masaharu Hayashi (Center for Information and Communication Technology,  
Hitotsubashi University)

NATSUME Takumi (University Library, Hitotsubashi University)

Kuninori Matsuda/ Kazuaki Yamamoto (Center for Collaborative Research on Pre-modern  
Books, National Institute of Japanese Literature)

Kanji Akagi (Faculty of Law, Keio University)

In this study, we describe a classification and arrangement method using web annotation model for Visual Material of Pre-modern Japanese Texts. Especially, we discuss to define a "Koda" virtual collection for Hitotsubashi University Library and Keio University Library and to clarify some issues in our practice.

### 1 はじめに

近年、大学等の研究機関が所蔵する資料画像に対して、Digital Object Identifier (DOI) に代表される永続的識別子を付与し、デジタルリポジトリ上で公開する事例が増加している[1-3]。資料画像に付与された永続的識別子により、論文等から資料画像を直接引用することが可能となり、論文等で研究利用した資料を明示することが可能となってきた。

ところが、こうした資料画像は、当然のことではあるが、所蔵機関の視点で整理・分類された形で公開されていることが多い。たとえば、異なる所蔵機関から同一作者の資料画像が異なる分類で公開される可能性も大いにある。そのため、こうした資料画像を利用する研究者は、ある程度の当たりをつけた上で、複数のデジタルリポジトリにアクセスし、目的の資料を探索しながら、独自の分類・整理を行うことになる。探索の過程で発見した資料画像は、研究者個人のローカル環境にその資料に関する情報とともに保存し、比較・分析を行いながら研究を進めることになる[4]。結果として、研究者のローカル環境には一種のコレ

クションができあがることになる。こうしたコレクションは同じ分野の研究者にとっては有益なコレクションとなる可能性があるが、基本的に公開されることはない。

一方でウェブ上のリソースや情報の整理・分類を目的としたウェブアノテーション技術がある。ウェブアノテーションとは注釈（アノテーション）の付与をウェブの世界で実現する手法・技術である。現在、W3C Web Annotation Working Group (WG) により、ウェブアノテーションのデータモデル Web Annotation Data Model [5] の標準化作業が進められている。このウェブアノテーション技術を用いれば、ウェブ上で公開されている資料画像の整理・分類が容易になる可能性がある。

そこで本研究では、慶應義塾大学図書館および一橋大学附属図書館が所蔵する「幸田文庫」に着目し、ウェブアノテーションを利用して仮想的なコレクション定義することを考えた。「幸田文庫」とは、日本近世経済史・日欧交渉史研究の分野で活躍した歴史学者、幸田成友（1873-1954）の旧蔵書で、現在、慶應義塾大学図書館および一橋大学附属図書館に分かれて所蔵されている。慶應義塾大学図書館所蔵の「幸田文庫」には、書誌学的

に貴重な春日版、五山版等が多数含まれている。一方、一橋大学附属図書館所蔵の「幸田文庫」には武鑑・町鑑等の実用書および幸田自身による書写資料が含まれていることが知られている[6]。近年、幸田成友の歴史研究の解明の解明が進められているが、幸田の研究範囲は経済誌研究、日欧通交史研究、書誌学などと広く、全容解明を難しくさせている[7]。現在、慶應義塾大学図書館の「幸田文庫」、一橋大学附属図書館の「幸田文庫」はそれぞれの機関リポジトリ上で部分的ではあるが公開状態にある。この二つの「幸田文庫」を一つの仮想コレクション「幸田文庫」として表現可能となれば、幸田研究の全容解明の一助となりうる。

本稿では、この二つの「幸田文庫」を対象にした、ウェブアノテーション技術による仮想コレクション定義手法を提案・実装し、その実現に向けた課題を明らかにする。

## 2 ウェブアノテーションによる仮想コレクション定義手法

本項では資料画像のためのウェブアノテーションモデルおよびウェブアノテーションを利用した仮想コレクション定義について述べる。

### 2.1 資料画像のためのアノテーションモデル

本研究では資料画像のためのアノテーションモデルを定義する。Web Annotation Data Modelをベースにしたアノテーションモデルである。

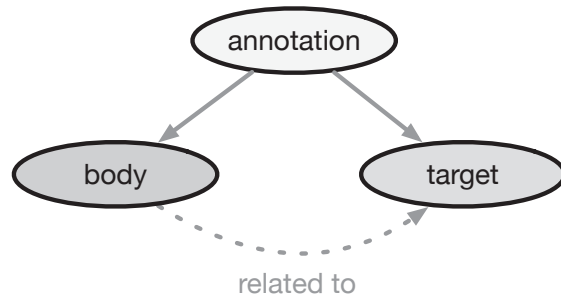


図1 Web Annotation Data Model

Web Annotation Data Modelはウェブリソースへの汎用的なアノテーションを表現するモデルである。アノテーションを表現する「Annotation」、アノテーション対象を表現する「Target」、アノテーション内容を表現する「Body」の3種類のウェブリソースから構成される(図1)。ウェブリソースの部分および全体に対してアノテーションを付与することができる。

これを資料画像へのアノテーションに応用したのが資料画像のためのアノテーションモデルである(図2)。Web Annotation Data Modelをベースに、資料画像の整理・分類に必要となる、アノテーションの作者、作成・更新日時、分類タグを保持する属性を追加している。とくに分類タグは資料画像の分類時の鍵となることを想定している。

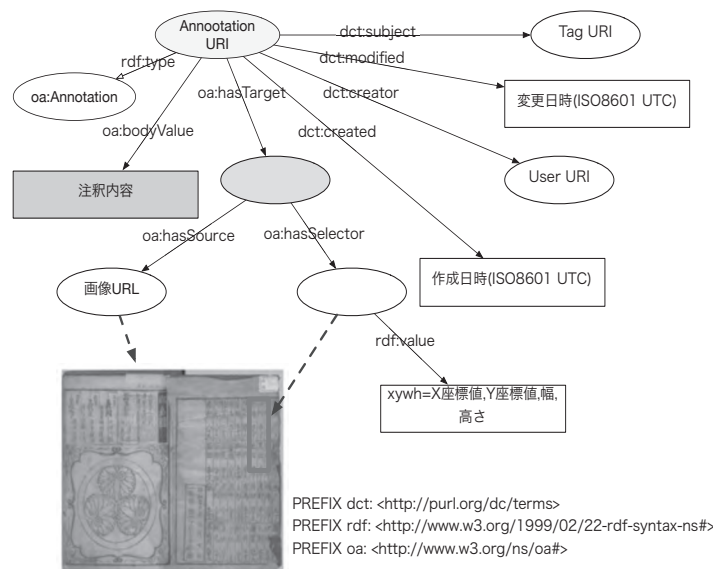


図2 資料画像のためのアノテーションモデル

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix oa: <http://www.w3.org/ns/oa#>.
@prefix dct: <http://purl.org/dc/terms/>.

<http://koda.nijl.ac.jp/annot/b5cb689e#xywh=431,256,377,224>
  rdf:type oa:Annotation ;
  dct:subject <http://koda.nijl.ac.jp/tag/幸田文庫> ;
  dct:creator <http://koda.nijl.ac.jp/who/user2> ;
  dct:created "2016-08-26T08:05:16Z" ;
  dct:modified "2016-08-26T08:05:16Z" ;
  oa:hasTarget _:b2286228124_arc48dbb1 ;
  oa:bodyValue "武鑑 #幸田文庫" .

_:b2286228124_arc48dbb1
  oa:hasSource <http://koda.nijl.ac.jp/img/KDqb00438040001.jpg> ;
  oa:hasSelector _:b290210598_arc48dbb2 .

_:b290210598_arc48dbb2
  rdf:value "xywh=431,256,377,224" .

```

リスト 1 アノテーション例 (Turtle 形式)

資料画像のためのアノテーションモデルによる記述例をリスト 1 に示す。この例では「oa:hasSource」が指し示すアノテーション対象 URL に対して「oa:hasSelector」が示す領域に「oa:bodyValue」で示される「武鑑 #幸田文庫」をアノテーションとして付与している。なお、本研究ではアノテーションモデルを Resource Description Framework (RDF) 形式で保存する。

## 2.2 仮想コレクション

仮想コレクション定義は資料画像を検索し、その資料画像へアノテーションを付与する形で行われる。リスト 1 の例では資料画像に付与されたアノテーションにより、資料画像にタグ「幸田文庫」が割り当てられている。仮想コレクションの定義はこうしたタグを付与し、検索する形で実現される。

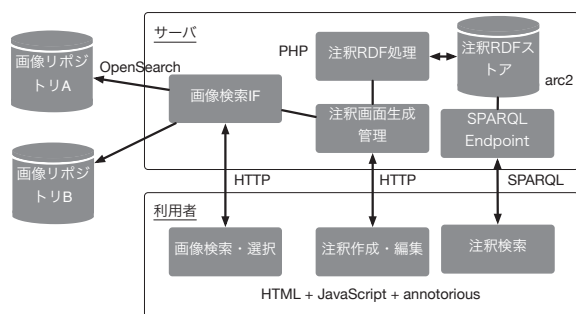


図 3 システム構成図

## 3 ウェブアノテーションシステム

本項では提案手法を実装したウェブアノテーションシステムおよび仮想コレクション提示手法について述べる。

### 3.1 システム概要

ウェブアノテーションシステムの全体構成を図 3 に示す。システムは画像検索インタフェース (IF)、アノテーション機能 (注釈画面生成管理、注釈 RDF 処理)、RDF ストア機能 (注釈 RDF ストア、SPARQL Endpoint) から構成される。

サーバ部の開発言語は PHP、利用者 IF は HTML と Javascript の組合せで開発した。アノテーション機能は Annotorious[8]、Portable Document Format (PDF) ファイルの処理に PDF.js[9]、画像ビューワは OpenSeadragon[10]を用いた。サーバ部の RDF 処理には ARC2[11]ライブラリを利用した。

### 3.2 画像検索機能

画像検索機能はプラグイン機構を備えており、拡張性を考慮した設計となっているが、本研究では OpenSearch プロトコル[12]を用いる。

OpenSearch は米アマゾンの子会社 A9 によって制定された、検索結果をフィードで提供するプロトコルである。大学等の機関が資料画像を公開する

ツールとして利用するデジタルリポジトリシステムでの実装が多いプロトコルである。

本研究では OpenSearch の検索結果である Atom フィードに資料画像 URL を含めることで資料画像検索を実現する。資料画像の URL は link 要素を利用して設定する (リスト 2 参照)。ウェブアノテーションシステムは link 要素の type 属性値を確認することで、アノテーション対象となる画像のファイルタイプを決定する。

### 3.3 アノテーション機能

アノテーション機能は検索結果から得られた資料画像に対してウェブアノテーションを実現するための機能を有する。実際のアノテーション付与画面を図 4 に示す。この画面では、資料画像全体または資料画像の部分に対するアノテーションの付与・編集・削除が可能である。アノテーション対象が PDF ファイルの場合は PDF ファイルの各ページを画像ファイルに変換し、ページ捲りができるように表示する。また、IIIF Presentation API [13] で記述された Manifest ファイルを開いた場合は PDF ファイルの場合と同様にページ捲りに対応した形で表示する。

```
<link rel="enclosure" title="
KDqb00438040001.jpg"
type="image/jpeg"
href=" http://koda.nijl.ac.jp/img/
KDqb00438040001.jpg "
/>
```

リスト 2 資料画像 URL の検索 (Opensearch feed の Link 要素)

### 3.4 RDF ストア機能

ウェブアノテーションシステムを介して資料画像に付与されたアノテーションは RDF 形式で RDF ストアに保存される。保存されたアノテーションは SPARQL クエリにより検索可能である。リスト 3 はアノテーションを検索する SPARQL クエリである。

```
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX dct: <http://purl.org/dc/terms/>

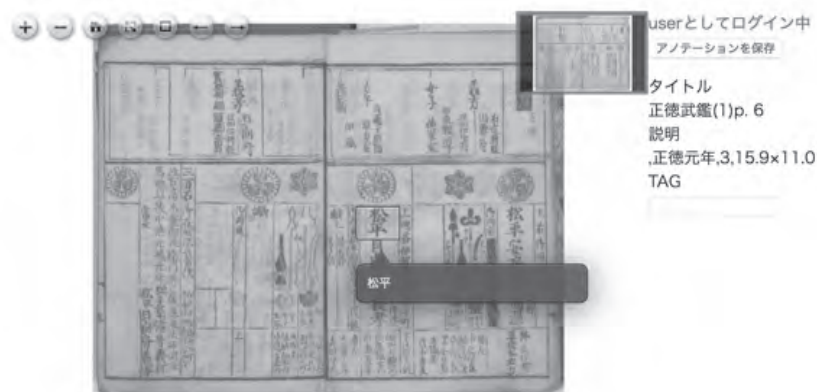
SELECT ?imguri WHERE {
  ?s dct:subject
  <http://koda.nijl.ac.jp/tag/幸田文庫> .
  ?s oa:hasTarget ?target .
  ?target oa:hasSource ?imguri .
}
```

リスト 3 ウェブアノテーション検索例 (SPARQL クエリ)

### 3.5 仮想コレクション定義

本研究では、ウェブアノテーションをキーに資料画像を収集することで資料画像群 (コレクション) を定義する。通常の画像ファイル (PDF ファイルを含む) の場合はリスト表示で仮想コレクションを表示する。アノテーション対象が IIIF Manifest の場合は IIIF Presentation API の manifests 要素を利用して IIIF Presentation API によるコレクション定義を行う (リスト 4 参照)。

## '正徳武鑑(1)'の注釈



[注釈トップページに戻る](#)

国文学研究資料館 古典籍画像アノテーションシステム

図 4 ウェブアノテーション付与画面例

```

{
  @context: "http://iiif.io/api/presentation/2/context.json",
  @id: "http://koda.nijl.ac.jp/iiif/collection/top",
  @type: "sc:Collection",
  label: "バーチャル幸田文庫",
  description: "泣き別れとなっていた幸田文庫が今一つに",
  attribution: "Provided by 慶應義塾図書館, 国文学研究資料館, 一橋大学附属図書館",
  collections: [ ],
  manifests: [
    {
      @id: "http://koda.nijl.ac.jp/iiif_test/10812535316.json",
      @type: "sc:Manifest",
      label: "本佐録 2巻. 乾(1).p. 1"
    },
    {
      @id: "http://koda.nijl.ac.jp/iiif_test/10812535324.json",
      @type: "sc:Manifest",
      label: "本佐録 2巻. 坤(1).p. 1"
    }
    ...省略...
  ]
}

```

リスト 4 仮想コレクション定義 (IIIF Presentation API)

このコレクション定義を IIIF 対応画像ビューワ Universal Viewer[14]等の IIIF 対応画像ビューワ (ただし manifests 要素を処理可能であること) で表示すると、一つのウィンドウで異なる機関の資料画像を閲覧することが可能となる。

## 4 考察

本項では、本研究におけるウェブアノテーションを利用した仮想コレクション定義手法について考察する。

### 4.1 資料画像の検索

提案手法であるウェブアノテーションシステムでは、アノテーションの付与前に資料画像の検索が必要である。本研究が対象とした「幸田文庫」は、一橋大学附属図書館、慶應義塾大学図書館の双方が機関リポジトリによる資料公開を実施していたため、OAI-PMH によるメタデータ収集、OpenSearch プロトコルの利用が比較的容易に実現できた。また、ある程度のメタデータがはじめから付与されていたことから検索可能であったという面もある。

しかしながら、必ずしも資料画像に関するメタデータが豊富に存在するとは限らない。検索によらない資料画像に対するアノテーション手法の検討も必要である。例えば、本研究では、ウェブ

ブラウザ拡張を利用したアノテーション付与手法の検討を始めている (図 5)。

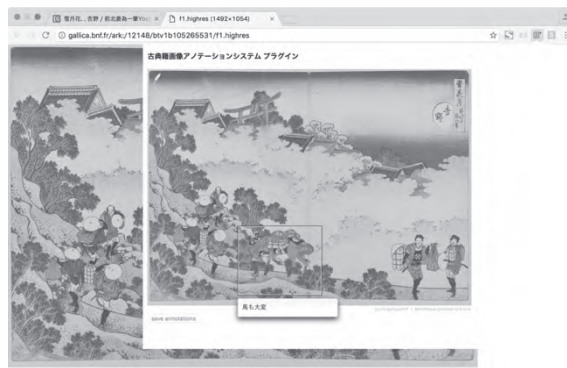


図 5 アノテーション付与画面 (ブラウザ拡張)

### 4.2 ウェブアノテーション付与

提案手法により、ウェブアノテーションを付与することで資料画像の分類・整理が可能となることは示すことができた。

しかしながら、本研究においては、アノテーション対象をウェブリソース以上のコンテキストで取り扱っており、そこが欠点でもある。すなわち、アノテーション対象が何であるかを、処理前に把握しておく必要があるということである。たとえば、アノテーション対象となる URL が指し示

すのは、資料全体なのかページ単位なのか、画像ファイルなのかPDFファイルなのか、または IIIF Manifest ファイルなのか、本手法によるアノテーション付与を実現するには、ウェブリソースのコンテキストを事前に把握しておかなければならない。この問題を解決するには、資料画像の表現手法を固定化するか、機械可読とするかといった対策が必要である。これらを鑑みると、資料画像のプレゼンテーション手法を定義する IIIF Presentation API の有用性は高いと考えている。

### 4.3 仮想コレクション定義

資料画像にウェブアノテーションを付与することで一種の資料画像群をウェブアノテーションで定義可能なことは示すことができた。しかしながら、仮想コレクション定義を行う際にも資料画像のプレゼンテーション手法を統一しておく必要があることから、この点においても IIIF Presentation API の有用性は高いと考えている。しかしながら、IIIF Presentation API では資料に関するメタデータ記述力が弱く、改善の余地がある。基本的なメタデータだけでなく、資料の物理的サイズや材質等、資料を識別可能な最低限必要な情報は含める必要がある。

## 5 まとめ

本研究では、ウェブアノテーションによる資料画像の統合および仮想コレクションの定義手法を提案した。統合対象の資料として、慶応義塾大学図書館および一橋大学附属図書館に別れて所蔵されている「幸田文庫」を対象に実験を行い、ウェブアノテーションを用いて画像資料の整理・分類が可能であること、仮想的なコレクションとして定義可能であることを示すことができた。しかしながら、対象に特化した面もあり、必ずしも汎用的な手法とはいえない。今後は資料画像の表現手法についても検討しながら、ウェブアノテーションによる仮想コレクション定義手法の確立を目指していきたい。

## 謝辞

本研究は、国文学研究資料館「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」における研究開発系共同研究「古典籍画像を対象としたメタデータ記述／アノテーションに関する研究」によるものである。また、一部技術の研究開発は科研費（基盤研究(B)：24300310、挑戦的萌芽研究：25560140、基盤研究(C)：15K00446）により行われた。最後に、慶應義塾大学図書館および一橋大学附属図書館の関係者、Xenon Limited

Partners の神崎正英氏にはこの場を借りて感謝の意を表す。

## 参考文献

1. 国立国会図書館デジタルコレクション, <http://dl.ndl.go.jp/> (参照 2016/10/01) .
2. 教育掛図資料リポジトリ, <https://wallchart.repon.org/> (参照 2016/10/01) .
3. OHSU Digital Commons, <http://digitalcollections.ohsu.edu/> (参照 2016/10/01) .
4. 大和 裕幸, 稗方 和夫, 畑野 勇, 新木 仁士: デジタルアーカイブを利用した歴史研究支援システムの構築, 情報処理学会研究報告, Vol. 2010-DD-74, No. 1, pp. 1-7 (2010) .
5. W3C: Web Annotation Data Model W3C Candidate Recommendation 05 July 2016. <https://www.w3.org/TR/annotation-model> (参照 2016/09/01) .
6. 高橋菜奈子: 幸田成友の経済史研究とその資料—一橋大学附属図書館所蔵幸田文庫を中心に—, 経済資料研究, Vol. 33, pp. 29-43 (2003) .
7. 夏目琢史: 幸田成友論—経済史から日欧通交史への断層—, 一橋大学附属図書館研究開発室年報, Vol. 4, pp. 18-51 (2016) .
8. Annotorious. <http://annotorious.github.io/> (参照 2016/10/01) .
9. PDF.js. <https://mozilla.github.io/pdf.js/> (参照 2016/10/01) .
10. OpenSeadragon. <https://openseadragon.github.io/> (参照 2016/10/01) .
11. ARC2. <https://github.com/semsol/arc2> (参照 2016/10/01) .
12. OpenSearch Specifications. [http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft\\_5](http://www.opensearch.org/Specifications/OpenSearch/1.1/Draft_5) (参照 2016/10/01) .
13. International Image Interoperability Framework (IIIF) Presentation API 2.1. <http://iiif.io/api/presentation/2.1/> (参照 2016/10/01) .
14. Universal Viewer. <https://github.com/UniversalViewer/universalviewer> (参照 2016/10/01) .