# IIIFとオープンデータを用いた『捃拾帖』内容検索システムの開発

中村 覚

東京大学
情報基盤センター



CC BY

# 発表内容

1. はじめに
2. データ作成
3. アプリケーション開発
4. 考察
5. 結論

# 1.はじめに

# 東京大学デジタルアーカイブズ構築事業

「東京大学ビジョン2020」

アクション１

③ 学術の多様性を支える基盤の強化

- *東京大学が保持する学術資産のアーカイブを構築し、その公開と活用を促進することで、学術の多様性を支える基盤を強化する*

# 東京大学デジタルアーカイブズ構築事業

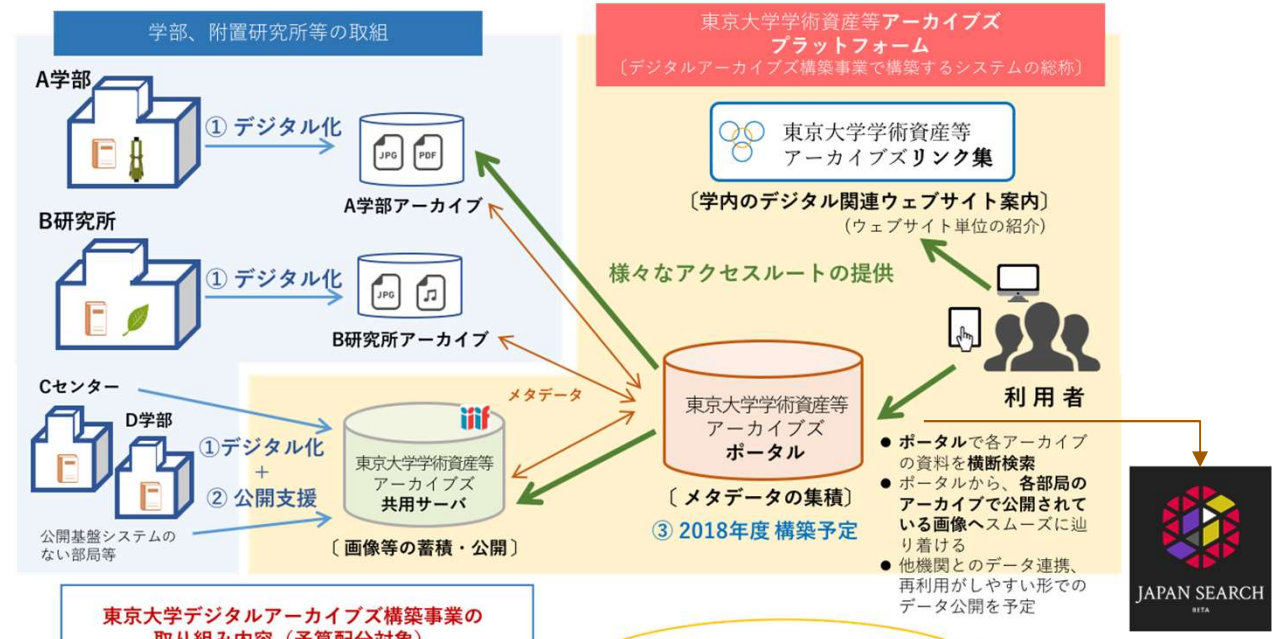
## 2016.09 東京大学学術資産等アーカイブズ委員会 設置

- 東京大学に学術資産等アーカイブズを構築し、及び活用することにより、学術の多様性を支える基盤の強化を図る

*東京大学学術資産等アーカイブズ委員会規則（平成28年9月29日 役員会議決 東大規則第17号）*

委員：(1) 附属図書館長
(2) 総合研究博物館長
(3) 文書館長
(4) 情報基盤センター長
(5) 総長が指定する部局の長から推薦された教授又は准教授 各１名
(6) その他総長が必要と認めた者

## 東京大学デジタルアーカイブズ構築事業の概要－公開支援（システム構成）について




**東京大学デジタルアーカイブズ構築事業の取り組み内容（予算配分対象）**

① 各部局での学術資産の**デジタル化促進**
② 画像等の**公開支援**
③ **ポータル**の**構築** ＝ メタデータの集約・発信

- **学術資産の発見可能性の向上**
- **データの利用・活用の促進**

相当

# 東京大学デジタルアーカイブズ構築事業

- 総合図書館
- 法学部
- 文学部
- 教育学部
- 情報学環
- 史料編纂所
- 文書館
- …

総合図書館『田中芳男・博物学コレクション』
https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/tanaka

東京大学文書館
デジタル・アーカイブ
https://uta.u-tokyo.ac.jp/uta

教育学部『富士川文庫』
https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/fujikawa

情報学環社会情報研究資料センター『Digital Cultural Heritage』
https://dch.iii.u-tokyo.ac.jp

Current Awareness Portal
図書館に関する情報ポータル

検索

English / 日本語

CA－R　CA－E　CA　調査研究

カレントアウェアネス・ポータルは、図書館界、図書館情報学に関する最新の情報をお知らせする、国立国会図書館のサイトです。

ホーム

# 東京大学総合図書館の公開するデジタル画像、利用申請なしで二次利用が可能に

Posted 2018年6月25日

2018年6月21日、東京大学附属図書館は、同学総合図書館が公開するデジタル画像が、利用目的を問わず利用申請なしで二次利用できるようになったことを発表しました。

今回対象となるのは、総合図書館の所蔵資料のうち、著作権の保護対象ではない資料のデジタル画像及びその関連データであり、かつインターネット上で公開されているものです。同館の所蔵資料であること、画像の改変を行った場合はその旨を明示すれば、利用申請なしで画像の転載や改変、再配布などの二次利用が可能です。

ニュース（東京大学附属図書館）
https://www.lib.u-tokyo.ac.jp/ja/library/contents/news
※「2018-06-21　総合図書館が公開するデジタル画像等が自由に使えるようになりました」とあります。

総合図書館が公開するデジタル画像等が自由に使えるようになりました（東京大学附属図書館）
https://www.lib.u-tokyo.ac.jp/ja/library/general/news/20180621

利用条件：画像データ等の利用について（東京大学附属図書館）
https://www.lib.u-tokyo.ac.jp/ja/library/general/reuse

**東日本大震災　関連情報**

「災害」に関する当サイトの記事

調査研究リポート「東日本大震災と図書館」

国立国会図書館　東日本大震災復興支援ページ

地震・災害 関連情報（レファレンス協同データベース）

各地の図書館等の被災情報等について（外部リンク）
saveMLAK

Twitterでの情報提供を行っています。
http://twitter.com/ca_tweet

## メニュー

- このサイトについて

東京大学学術資産等
アーカイブズ
共用サーバ

画像データの相互
利用の支援

Open
License

コンテンツの二次
利用の促進

Linked
Open Data

OSSの利用と
APIの提供

国際的・標準
的な規格・技
術の採用

テキストデータ
の共有と保存

https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/tanaka

# 田中芳男・博物学コレクション：2018年8月6日公開

# 課題：貼り込み資料単位で検索することができない

https://wwwap.hi.u-tokyo.ac.jp/ships/shipscontroller

# 東京大学史料編纂所：摺物データベース

1999年4月1日に公開

摺物データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ

検索結果：2645件　検索式：キーワード='(東京大学総合)'

ソート

1 2 3 4 5 6 7 8 9 次へ 最終へ

| No | 詳細 | 和暦年月日 | タイトル | 所蔵 |
|---|---|---|---|---|
| 1 | 詳細 | 安政５年 | 白人市引札 | 東京大学総合 6010-1-1 |
| 2 | 詳細 | 安政５年 | 〔米国１セント貨摺取図〕 | 東京大学総合 6010-1-10 |
| 3 | 詳細 | 安政５年 | 〔三河吉田本町合羽屋製なつとう引札〕 | 東京大学総合 6010-1-100 |
| 4 | 詳細 | 安政６年４月６日 | 〔名古屋桑名町２丁目知多郡酒売捌所加見屋常三郎店開店口演〕 | 東京大学総合 6010-1-101 |
| 5 | 詳細 | 安政５年 | 〔江戸神田明神前通坂下三ツ角あさだ長ゑもんあさだもぐさ引札〕 | 東京大学総合 6010-1-102 |
| 6 | 詳細 | 安政５年 | 〔名古屋本町５丁目京屋清統菓子司引札〕 | 東京大学総合 6010-1-103 |
| 7 | 詳細 | 安政５年 | 〔名古屋本町１２丁目丸屋佐右衛門砂糖類引札〕 | 東京大学総合図書館-A00-6010-1-104 |

摺物データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ

前　件　次　件　先　頭　最　終　一　覧

1/2645件

| | |
|---|---|
| 【和暦年月日】 | 安政５年(18580550550) |
| 【タイトル】 | 白人市引札 |
| 【版種類】 | 木版 |
| 【数量】 | １枚 |
| 【書誌事項】 | 19.5×14.5　「現金亭正味、手附庵三日、松乃舎」とあり |
| 【形態分類】 | 引札 |
| 【内容分類】 | 文芸・芸能・スポーツ・教育・出版・教化 |
| 【地名】 | 尾張？ |
| 【所蔵機関】 | 東京大学総合図書館-A00-6010-1-1 |
| 【目録番号】 | 16-A00-6010-1-1 |

# 目的

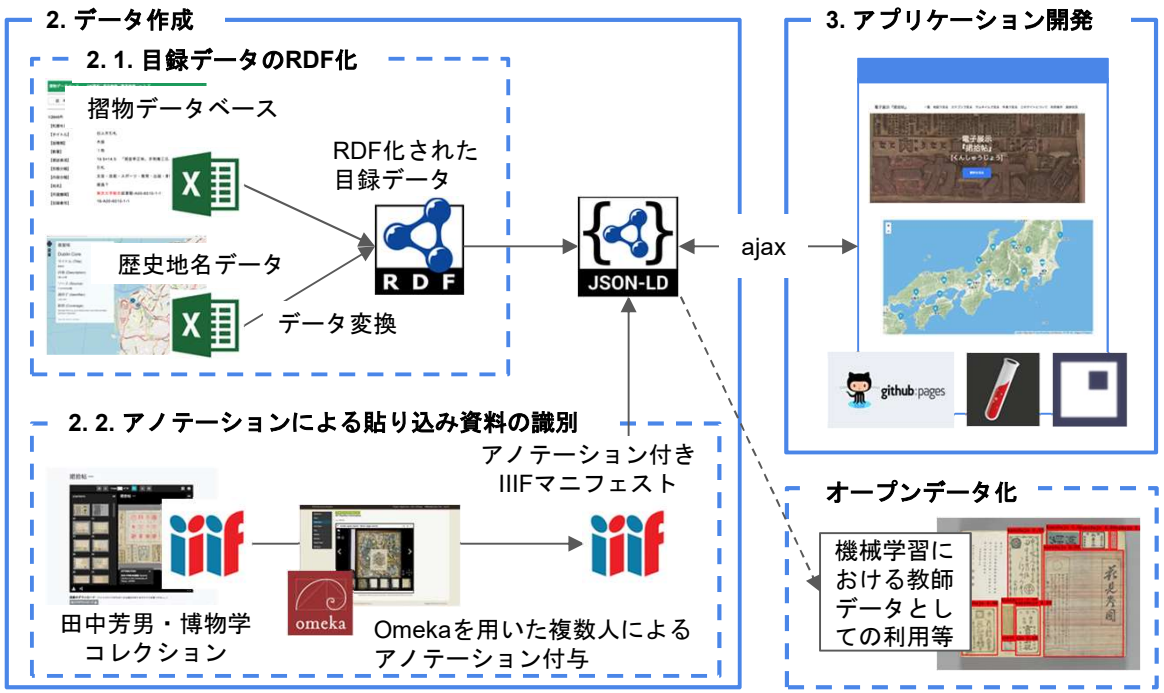
- 『裙拾帖』の内容（貼り込まれた資料）に基づく検索が可能なシステムを開発する。
  - IIIFの利点を活かした画像の部分切り出し

- オープンデータの利活用による多角的な資料提供方法を検討する。
  - 東京大学デジタルアーカイブズ構築事業：田中芳男・博物学コレクション
  - 東京大学史料編纂所：摺物データベース
  - 人文学オープンデータ共同利用センター：IIIF Curation Platform
  - 人間文化研究機構：歴史地名データ

http://codh.rois.ac.jp/face/

# 関連研究：顔貌コレクション

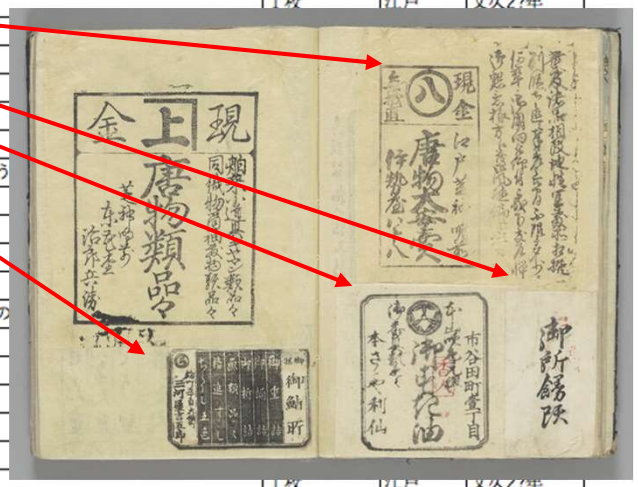
“「顔貌コレクション（顔コレ）」では、国文学研究資料館、慶應義塾大学、京都大学附属図書館が公開するIIIF画像から、顔貌だけを横断的に収集して新しいキュレーションを作成、検索できます。公開サイトの壁を乗り越えることで、美術史研究が進展することを期待しています。”
https://twitter.com/rois_codh/status/999167174043299840


2. データ作成
2. 1. 目録データのRDF化
摺物データベース
RDF化された
目録データ
歴史地名データ
データ変換
R D F
JSON-LD
ajax
3. アプリケーション開発
github pages
2. 2. アノテーションによる貼り込み資料の識別
アノテーション付き
IIIFマニフェスト
田中芳男・博物学
コレクション
omeka
Omekaを用いた複数人による
アノテーション付与
オープンデータ化
機械学習に
おける教師
データとし
ての利用等

# 2. データ作成

# 2. 1 目録データのRDF化：摺物データベース

| 9内容分 | 3西暦コード | タイトル | 7書誌事項 | 数量 | 地名 | 和暦 |
|---|---|---|---|---|---|---|
| 諸営業 | 18620199099 | 〔芝神明前伊 | 14.6×13.3 | 1枚 | 江戸 | 文久2?年 |
| 諸営業 | 18620199099 | 御所饅頭引札 | 9.5×5.2 | | | |
| 諸営業 | 18620199099 | 〔市谷田町１ | 10.0×9.5 | | | |
| 諸営業 | 18620199099 | 〔芝神明前東 | 18.0×14.0 | | | |
| 諸営業 | 18620199099 | 〔麹町５丁目 | 5.7×8.2 | | | |
| 諸営業 | 18620199099 | 〔上毛安中並 | 15.8×20.0 | | | |
| 諸営業 | 18620199099 | 〔京橋南伝馬 | 9.7×15.3　その他にいんげん豆、ぶどう | | | |
| 諸営業 | 18620199099 | ゆへし引札 | 22.0×11.0　水月堂とあり | | | |
| 諸営業 | 18620199099 | 御茶ふしの粉 | 7.8×4.6　通りしほ町大和屋小兵衛 | | | |
| 諸営業 | 18620199099 | 〔鉄砲州船松 | 15.3×23.5 | | | |
| 諸営業 | 18620199099 | 〔某品引札〕 | 8.8×4.1 | | | |
| 常識・娯楽 | 18620199099 | 浮世たとえ | 8.5×10.3　「男は当てくだけ」等１３の | | | |
| 諸営業 | 18620199099 | 御菓子引札 | 7.6×2.9 | | | |
| 諸営業 | 18620199099 | 御蒸菓子引札 | 7.7×3.5 | | | |
| 諸営業 | 18620199099 | 〔□□田町菊 | 3.2×2.4 | | | |
| 諸営業 | 18620199099 | 〔筋違御門外 | 2.5×3.0 | | | |
| 諸営業 | 18620199099 | 元祖御膳甘露 | 5.2×4.2　蒲田山本梅やしき | | | |
| 諸営業 | 18620199099 | 〔中橋鞘町角 | 4.0×3.0 | 1枚 | 江戸 | 文久2?年 |
| 諸営業 | 18620199099 | 薄ゆき引札 | 6.6×3.8 | 1枚 | 江戸 | 文久2?年 |

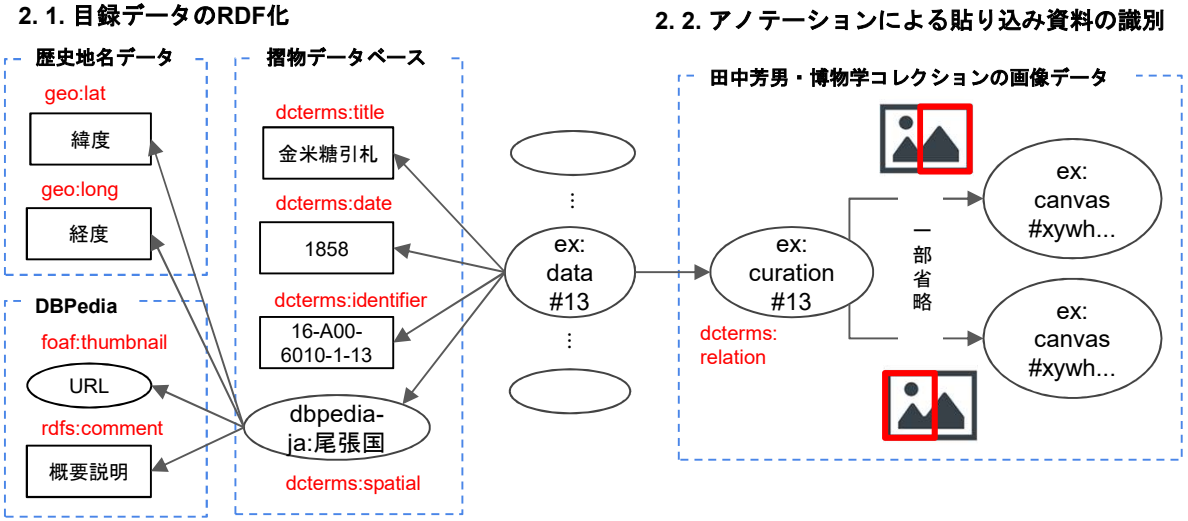


人間文化研究機構. 歴史地名データ. https://www.nihu.jp/ja/publication/source_map

# 2. 1 目録データのRDF化：歴史地名データ

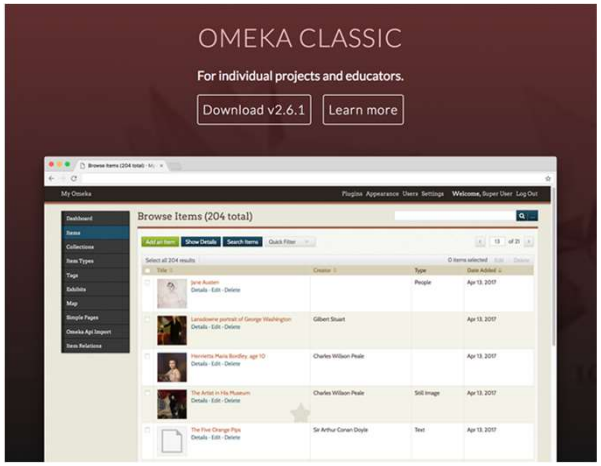
大日本地名辞書, 延喜式神名帳, 旧5万分の1地形図から地名を抽出し、その場所を現代もしくは過去の地図を用いて比定した上で、緯度・経度を求めたもの

# 2. データ作成


2. 1. 目録データのRDF化
歴史地名データ
geo:lat
緯度
geo:long
経度
DBPedia
foaf:thumbnail
URL
rdfs:comment
概要説明
摺物データベース
dcterms:title
金米糖引札
dcterms:date
1858
dcterms:identifier
16-A00-6010-1-13
dbpedia-ja:尾張国
dcterms:spatial
ex: data #13
2. 2. アノテーションによる貼り込み資料の識別
田中芳男・博物学コレクションの画像データ
ex: curation #13
dcterms: relation
一部省略
ex: canvas #xywh...
ex: canvas #xywh...

# 2. 2 アノテーションによる貼り込み資料の識別

- Omeka Classic
  - オンラインのデジタルコレクションのためのオープンソースのコンテンツ管理システム
  - テーマとプラグインで機能を拡張することができる
  - 人文学研究に適した様々なプラグインが数多く開発されている
- Plugin: IIIF Toolkit
  - 公開されているIIIFコンテンツを収集・共同編集するツール




- Omeka Classic, https://omeka.org/classic/
- IIIF Toolkit, https://omeka.org/classic/plugins/liifItems/

# 2. 2 アノテーションによる貼り込み資料の識別

①IIIFマニフェストのインポート

②メタデータの付与

③アノテーションの付与

# 2. 2 アノテーションによる貼り込み資料の識別

電子展示『珺拾帖』

一覧　地図で見る　カテゴリで見る　サムネイルで見る　年表で見る　このサイトについて　利用条件　進捗状況

アノテーション付与プロジェクト

Progress

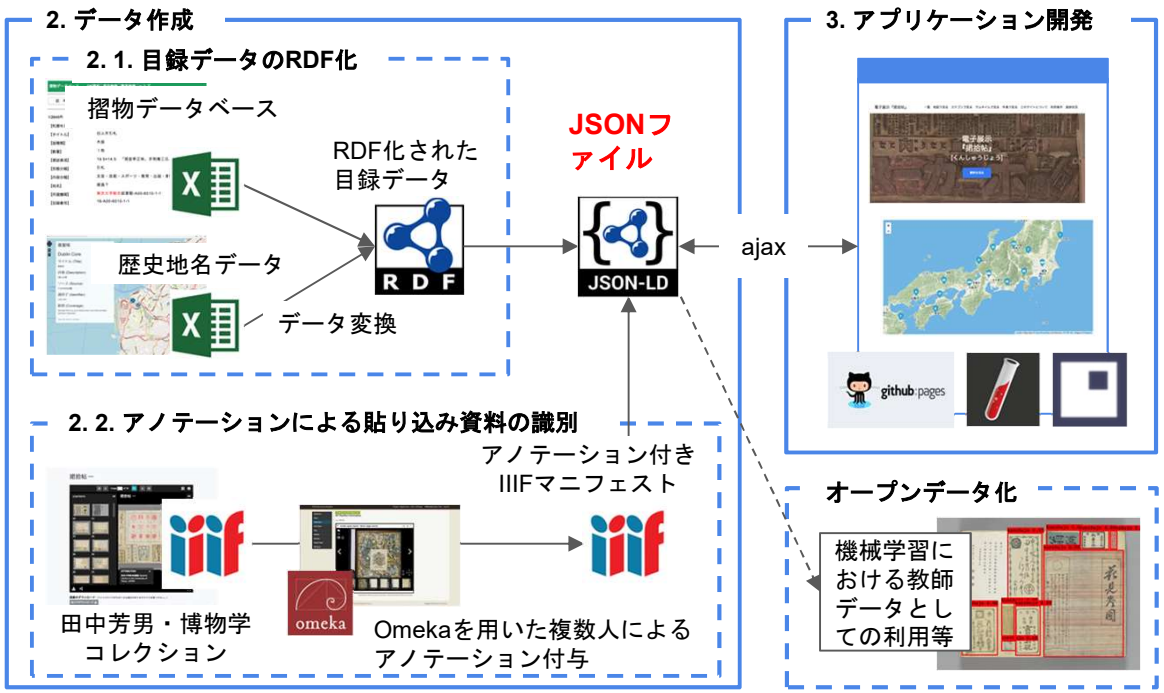
92 / 205

Show 10 entries　　Search:

| タイトル | 目録番号 | 書誌事項 | | その他 | 頁 |
|---|---|---|---|---|---|
| 「増補改正訳鍵PQRST四」題箋 | 16-A00-6010-5-1 | 17.5×3.2 | | 1 枚; 1810; 木版; | 2 |
| 「太政官日誌」題箋 | 16-A00-6010-5-2 | 14.3×3.5 | | 1 枚; 1868; 木版; | 2 |
| 「植学啓原三」題箋 | 16-A00-6010-5-3 | 17.4×3.0 | | 1 枚; 1834; 木版; | 2 |
| 〔新聞番附〕 | 16-A00-6010-5-4 | 24.5×16.0 | 関４月までの刊行新聞名あり | 1 枚; 1868; 木版; 番附; | 2 |
| 〔柳河春三・医学所・大病院地図〕 | 16-A00-6010-5-5 | 24.5×14.0 | | 1 枚; 1868; 木版; | 3 |
| 宝船図 | 16-A00-6010-5-6 | 17.5×18.0 | 「六十一羅漢」とあり | 1 枚; 1862; 木版; | 3 |
| 〔狐３匹の図〕 | 16-A00-6010-5-7 | 9.2×14.7 | | 1 枚; 1862; 木版; | 3 |
| 〔文久三年大小暦〕 | 16-A00-6010-5-8 | 12.5×18.0 | 玄魚画賛、「正月やゆつたり四霜うしろ六九」、獅子舞の色刷り、９とペアか | 1 枚; 1863; 木版; 暦; | 4 |
| 文久三年みつのとのゐ花暦 | 16-A00-6010-5-9 | 12.5×18.0 | ８とペアか | 1 枚; 1863; 木版; 暦; | 4 |
| 〔文久二年大小暦〕 | 16-A00-6010-5-10 | 17.1×9.0 | | 1 枚; 1862; 木版; 暦; | 4 |

Showing 1 to 10 of 205 entries　　Previous　1　2　3　4　5　...　21　Next

omeka


2. データ作成
2. 1. 目録データのRDF化
摺物データベース
RDF化された
目録データ
歴史地名データ
データ変換
R D F
JSONフ
ァイル
JSON-LD
ajax
3. アプリケーション開発
github pages
2. 2. アノテーションによる貼り込み資料の識別
アノテーション付き
IIIFマニフェスト
omeka
田中芳男・博物学
コレクション
Omekaを用いた複数人による
アノテーション付与
オープンデータ化
機械学習に
おける教師
データとし
ての利用等

# 3. アプリケーション開発

# 3. アプリケーション開発

- 静的サイトとして構築
  - Jekyll：静的サイトジェネレータ
  - GitHub Pages：静的サイトのホスティングサービス
- 理由
  - 検索対象とするデータ数が3,000件弱と小規模な点
    - 2,713件の貼り込み資料
    - https://kunshujo.dl.itc.u-tokyo.ac.jp/data/data.json
  - 運用コストの低減

# 電子展示『捃拾帖』[くんしゅうじょう]

https://kunshujo.dl.itc.u-tokyo.ac.jp/

電子展示『捃拾帖』

一覧　地図で見る　カテゴリで見る　サムネイルで見る　年表で見る　このサイトについて　利用条件　その他

選択した資料
を同一画面上で比較

選択行を比較　すべての選択を解除

2,645 件中 1 から 50 まで表示　1 row selected　50 件表示　検索:

前　1　2　3　4　5　…　53　次

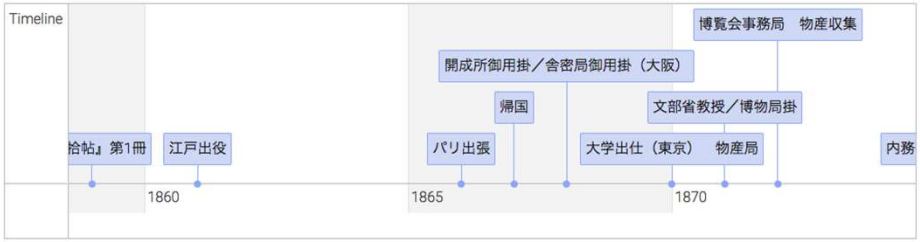
| | タイトル | 帖数 | 目録番号 | 書誌事項 | 枚数 | 年月日 | 地名 | 版 | 形態 | 内容 | ICP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | 白人市引札 | 『捃拾帖』第1帖 | 16-A00-6010-1-1 | 19.5×14.5　「現金亭正味、手附庵三日、松乃舎」とあり | 1枚 | 安政5年 | 尾張 | 木版 | 引札 | 文芸・芸能・スポーツ・教育・出版・教化 | |
| ☐ | 嘉永校定東西地球万国全図 | 『捃拾帖』第1帖 | 16-A00-6010-1-2 | 22.0×15.5　東都文渓堂梓、栗原信晁再校、為斎画、但し包紙のみ、オランダ語でKAARTとあり | 1枚 | 嘉永6年 | 江戸 | 木版 | | 常識・娯楽・遊戯・地図・食事 | |
| ☐ | 芽子垣内月並兼題 | 『捃拾帖』第1帖 | 16-A00-6010-1-3 | 15.5×20.7　1月は春従東来とこうばい、12月は炭竈焔細とひかげ | 1枚 | 安政6年1月 | 尾張 | 木版 | | 文芸・芸能・スポーツ・教育・出版・教化 | |

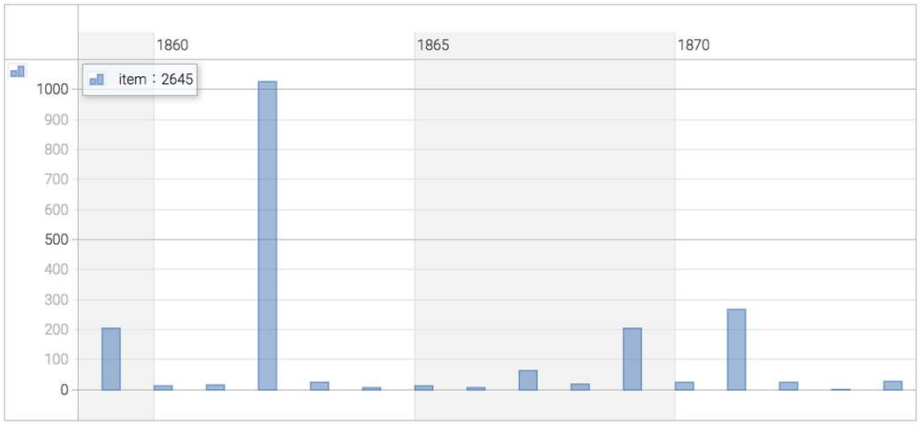
IIIF Curation Playerで閲覧

# 地図で見る

確認
「尾張国」を検索しますか？
尾張国（おわりのくに/をはりのくに）は、かつて日本の地方行政区分だった令制国の一つ。東海道に属する。 ウィキペディア
閉じる
検索

# 年表
# で見る




ヒストグラムをダブルクリックで詳細を表示

# 4. 考察

# オープンデータの利活用（マッシュアップ）

- （IIIF）画像データ
  - 東京大学デジタルアーカイブズ構築事業：田中芳男・博物学コレクション
- 「捃拾帖」貼り込み資料単位のメタデータ
  - 東京大学史料編纂所・摺物データベース
- 歴史地名データ
  - 人間文化研究機構

⬇

- 電子展示『捃拾帖』：内容検索システムの開発

# オープンソースソフトウェアの利用

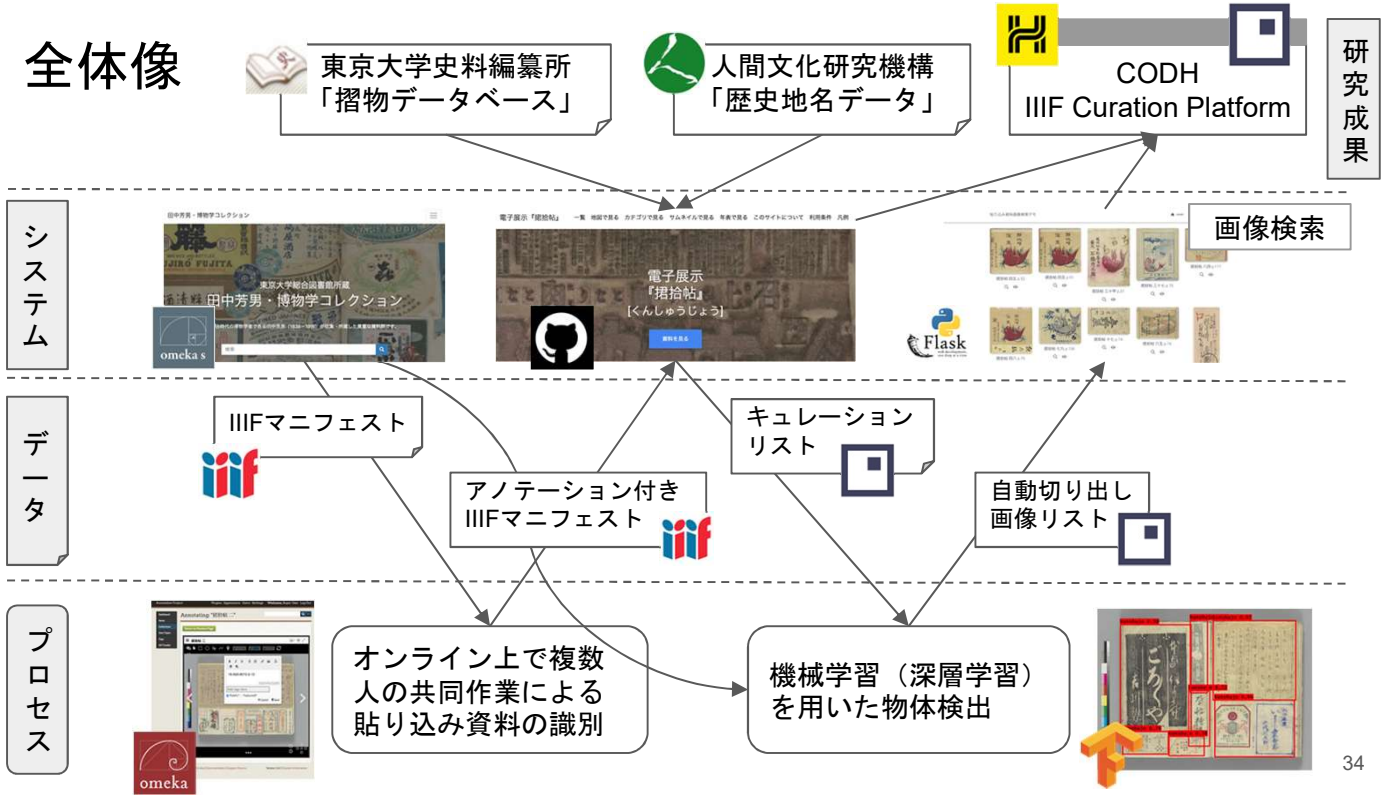
- Omeka + プラグイン「IIIF Toolkit」
  - （IIIF）画像の領域指定・切り出し
- 人文学オープンデータ共同利用センター・IIIF Curation Platform（Player）
  - 切り出したIIIF画像（貼り込み資料）の表示
- Mirador、Web地図ライブラリ「Leaflet」、etc...
- Omeka S
  - デジタルコレクションの構築とIIIF画像公開

# 今後の展望

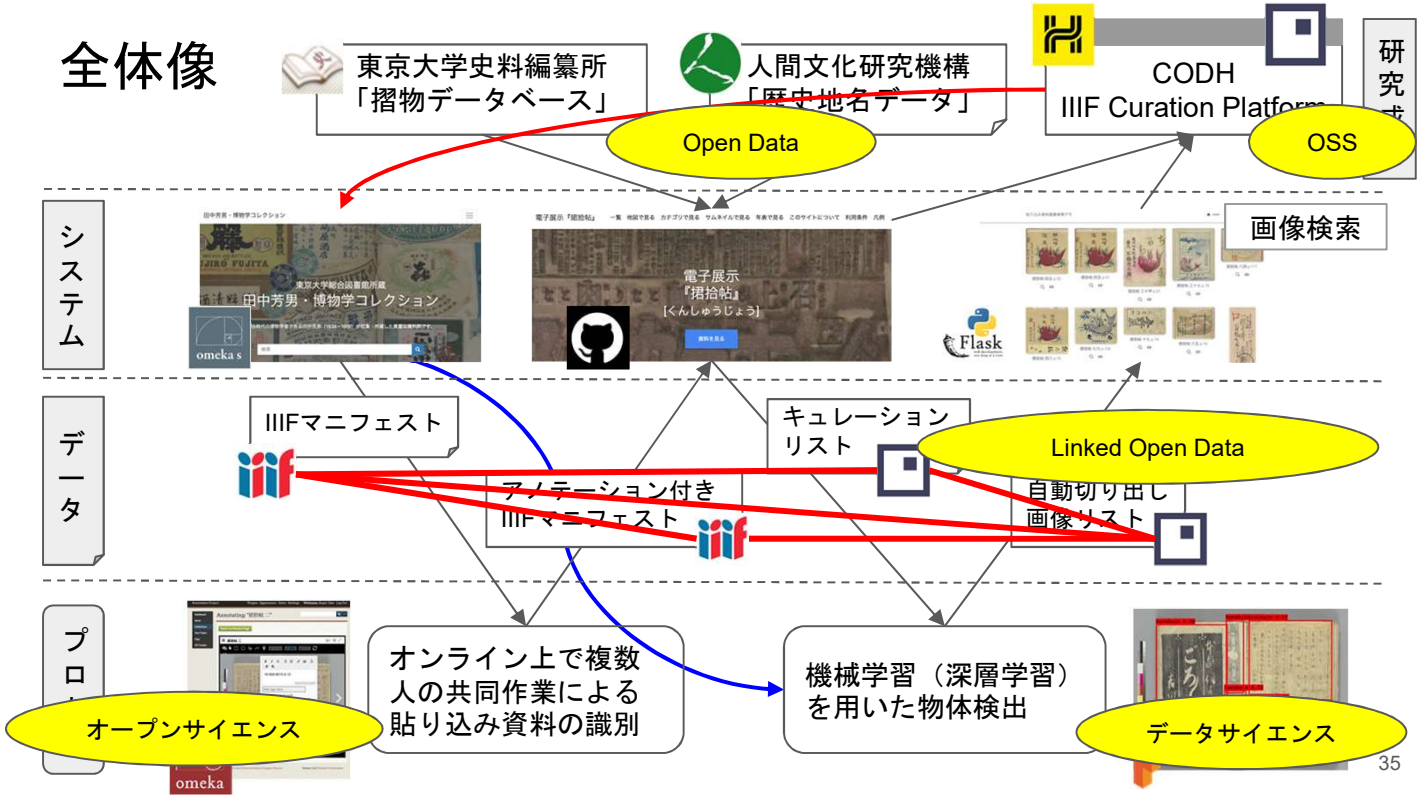
https://kunshujo.dl.itc.u-tokyo.ac.jp/data/data.json

- 機械学習を利用した貼り込み資料の自動切り出し：本研究において人手で作成した切り出し画像を教師データとして利用する
  - 切り出し済み帖数：15帖（第1～15帖）
  - 残帖数：約90帖（『外国捃拾帖』等の他のシリーズも含む）
- 切り出し画像に関する情報のオープンデータ化（JSON-LD形式による提供）
  - 第三者によるデータ利用、共同研究の可能性

# 全体像


東京大学史料編纂所
「摺物データベース」
人間文化研究機構
「歴史地名データ」
CODH
IIIF Curation Platform
研究成果
システム
田中芳男・博物学コレクション
omeka s
電子展示
『摺拾帖』
[くんしゅうじょう]
Flask
画像検索
データ
IIIFマニフェスト
アノテーション付き
IIIFマニフェスト
キュレーション
リスト
自動切り出し
画像リスト
プロセス
omeka
オンライン上で複数
人の共同作業による
貼り込み資料の識別
機械学習（深層学習）
を用いた物体検出

# 全体像


東京大学史料編纂所
「摺物データベース」
人間文化研究機構
「歴史地名データ」
CODH
IIIF Curation Platform
研究
Open Data
OSS
システム
画像検索
田中芳男・博物学コレクション
電子展示
『摺拾帖』
[くんしゅうじょう]
omeka s
Flask
データ
IIIFマニフェスト
キュレーション
リスト
Linked Open Data
アノテーション付き
IIIFマニフェスト
自動切り出し
画像リスト
プロ
オンライン上で複数
人の共同作業による
貼り込み資料の識別
機械学習（深層学習）
を用いた物体検出
オープンサイエンス
データサイエンス
omeka

# 5. 結論

# 結論

- 『捃拾帖(くんしゅうじょう）』の内容検索を可能とするシステムの開発事例について述べた。

- 複数の機関が提供する各種リソース（IIIF準拠画像・OSS・オープンデータ）を組み合わせて利用（マッシュアップ）した点を特徴とする。
  - デジタルアーカイブの利活用を検討する上での一事例として

関係者・機関に深く感謝いたします。

- 東京電気通信大学　佐藤賢一先生
- 東京大学史料編纂所　横山伊徳先生、鶴田啓先生、高橋敏子先生
- 東京大学学術資産アーカイブ化推進室

- 人文学オープンデータ共同利用センター
- 人間文化研究機構

ご静聴ありがとうございました。