

カルチャーデータとデジタルヒューマニティーズ ～くずし字認識、日本美術、歴史ビッグデータ～

北本 朝展

ROIS-DS人文学オープンデータ共同利用センター（CODH）

国立情報学研究所

<https://researchmap.jp/kitamoto/>

kitamoto@nii.ac.jp

自己紹介

<https://researchmap.jp/kitamoto/>



@KitamotoAsanobu

- 北本朝展（きたもと あさのぶ）
- ROIS-DS人文学オープンデータ
共同利用センター センター長
- 国立情報学研究所 教授
- 1997年 東京大学大学院工学系
研究科電子工学専攻修了／博士
（工学）
- 専門は情報工学、デジタル・
ヒューマニティーズ、データ駆
動型サイエンス（地球科学・防
災等）等



デジタル台風：台風画像と台風情報
<http://agora.ex.nii.ac.jp/digital-typhoon/>

はじめに

ROIS-DS人文学 オープンデータ 共同利用セン ター（CODH）

<http://codh.rois.ac.jp/>



2016年4月～

「オープン」の概念を核として三者
を接続し、知識の深化、巨大化、多
様化を目指す

深化

研究者

巨大化

機械



市民

メンバー
国立情報学研究所
統計数理研究所
センター長＋
特任助教4名

多様化

人文学におけるデジタル変革

データ駆動型人文学

1. データを基盤とする新しい手法により、**人文学の研究手法を革新**する。
2. **人文学的に新しい知識**の獲得を目標とする。
3. データの共有やエビデンスに基づく検証などの**新しい文化**を人文学に導入する。

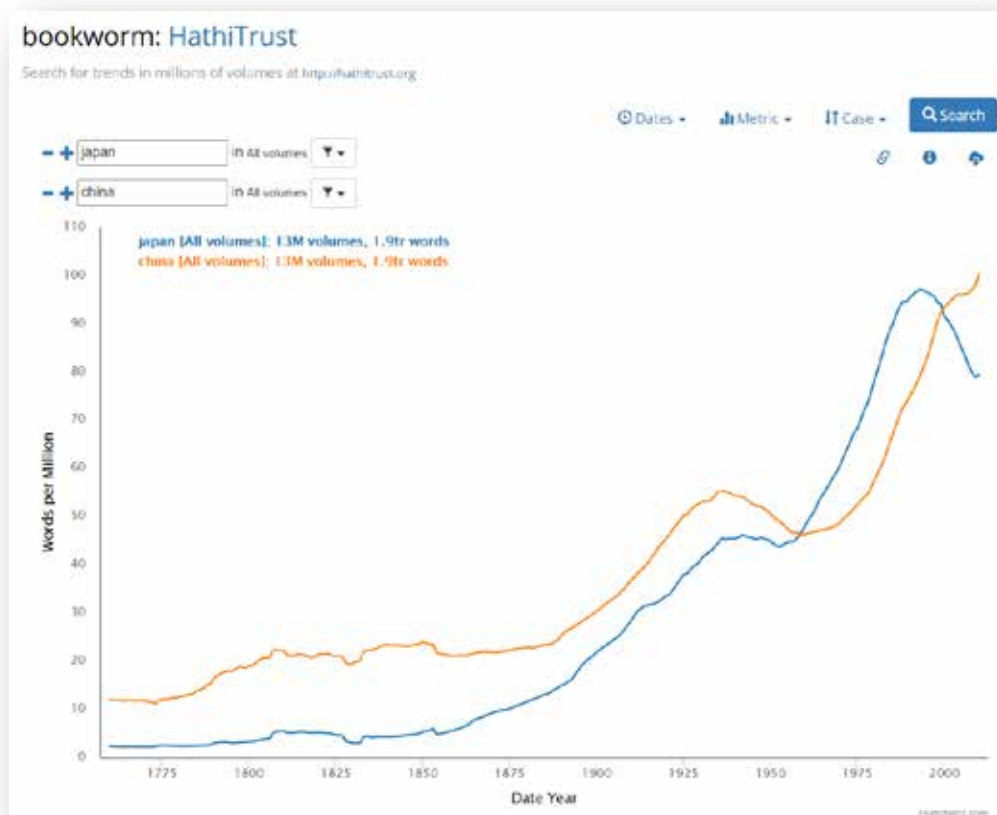
人文学ビッグデータ

1. 人文学が生産する大規模データにより、**(非)人文学の研究手法を革新**する。
2. **データに基づく(定量的)知識**の獲得を目標とする。
3. 人文学データに特有の構造化や解釈などの**技術を分野を超えて共有**する。

デジタルヒューマニティーズ

1. 人文学において、デジタル技術に基づき研究方法を革新する分野。
2. 日本語では「人文情報学」「デジタル人文学」とも言われる。
3. デジタルデータやツールの整備には、優秀な人材と新たな研究資金が必要。
4. 情報学者と人文学者が相互に恩恵を受けるように研究計画を立案する。

ビッグデータと検索

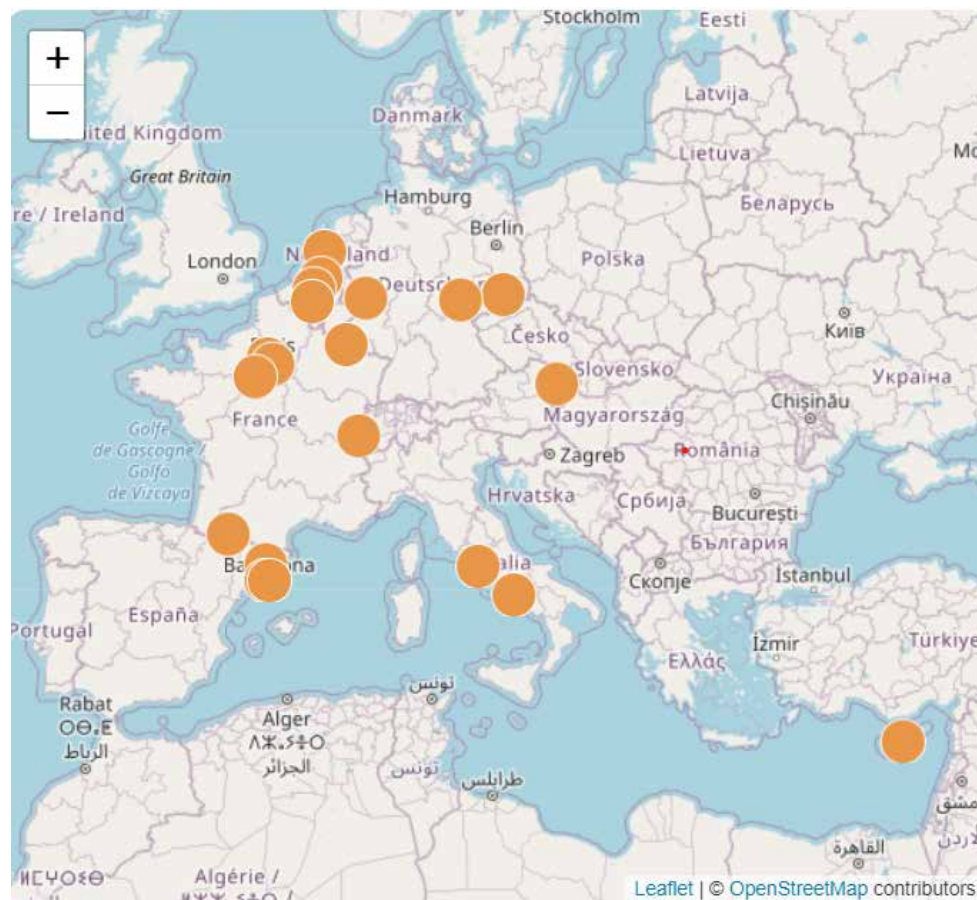


HathiTrust bookworm,
<https://bookworm.htrc.illinois.edu/develop/>

1. 人文学ビッグデータに基づく語（N-gram）の出現頻度の推移。Google Books Ngram Viewer や HathiTrust bookwormなどで検索可能。
2. 検索機能が人文学研究のスピードとスタイルを変える。
3. OCRが困難なため、日本語では同様の検索は困難。

Time Machine Europe

<https://www.timemachine.eu/>



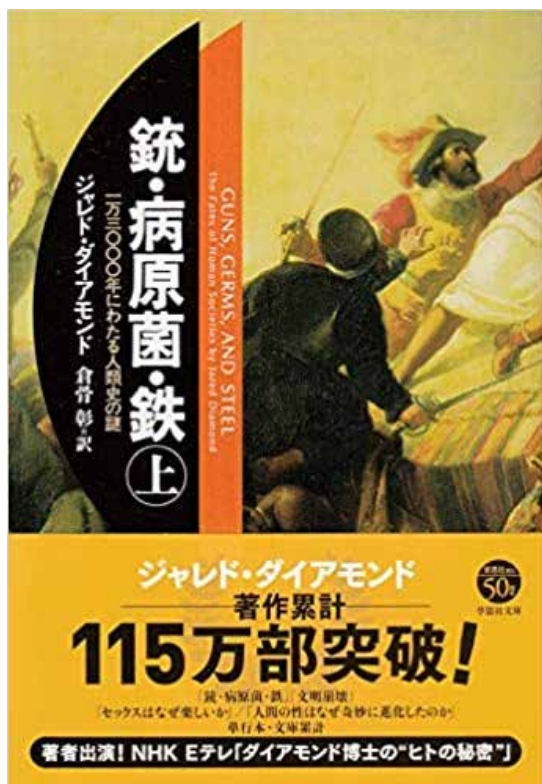
- 欧州の文化的資産をデジタル化・構造化し、誰もがアクセスできる「過去のビッグデータ（Big Data the Past）」を構築。
- 革新的なデジタル化技術や人工知能（AI）技術などを開発。
- 欧州委員会のフラグシップ研究計画 = 10年で1000億円（見直し中）。参加機関700以上。

<https://current.ndl.go.jp/e2248>

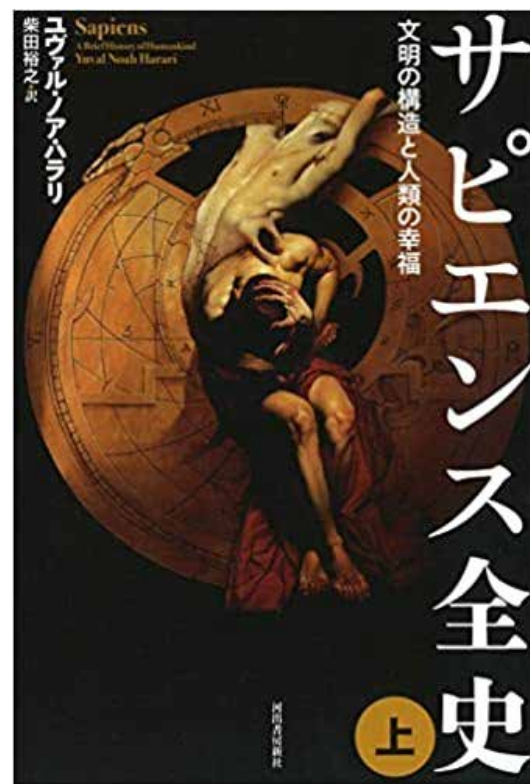
LTM Projects, <https://www.timemachine.eu/ltm-projects/>

グローバル／ビッグヒストリー

「歴史」の
対象が空間的・社会的
に拡大し、
より分野横
断的になっ
てきた



ジャレド・ダイヤモンド、銃・病原菌・鉄



ユ瓦尔・ノア・ハラリ、サピエンス全史



デヴィッド・クリスチャン、ビッグヒストリー われわれはどこから来て、どこへ行くのか——宇宙開闢から138億年の「人間」史

くずし字認識

共同研究者：カラーヌワット タリン（CODH）、Alex Lamb（モントリオール大）、Mikel Bober-Irizar（ケンブリッジ大）

くずし字データセットの利活用例

<http://codh.rois.ac.jp/char-shape/>

日本古典籍
データセット
(国文研蔵)

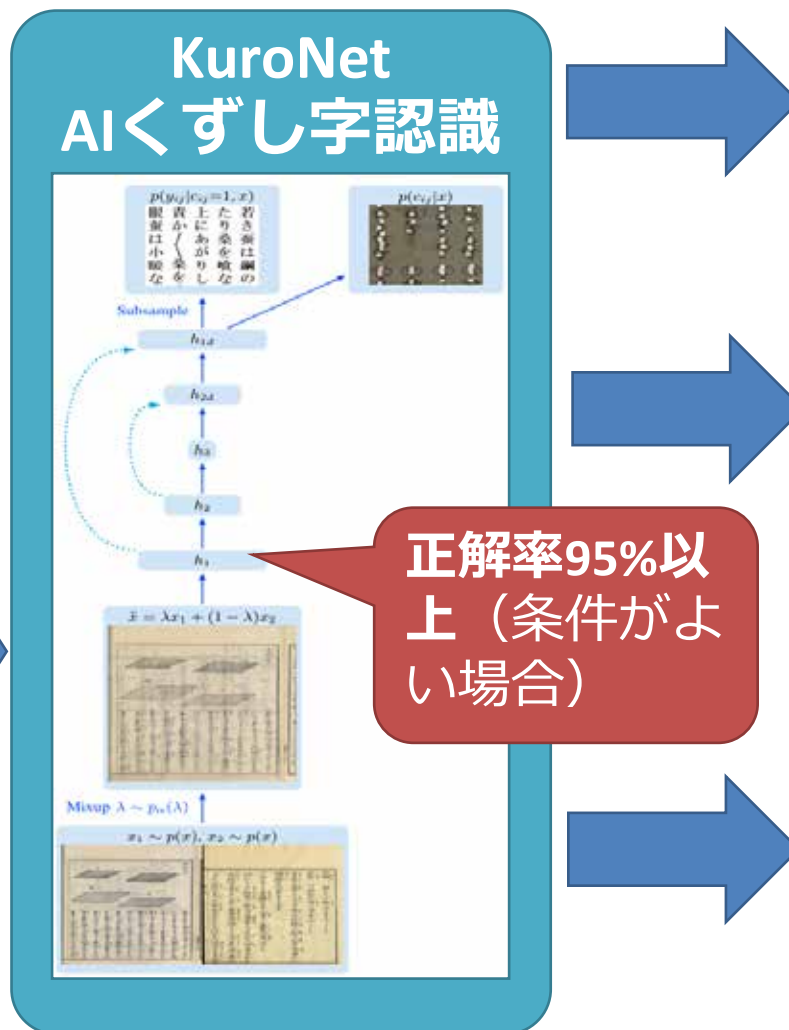


くずし字データ
セット (国文研・
CODH作成)



file	char	x	y
200003803_00024_2.jpg	U+3067	416	114
200003804_00024_2.jpg	U+3055	232	115
200003805_00024_2.jpg	U+304A	327	115
200003806_00024_2.jpg	U+3068	145	116
200003807_00024_2.jpg	U+3046	369	116
200003808_00024_2.jpg	U+305F	457	116
200003809_00024_2.jpg	U+5FA1	104	117
200003810_00024_2.jpg	U+3072	191	118
200003811_00024_2.jpg	U+540D	279	120
200003812_00024_2.jpg	U+3061	501	120

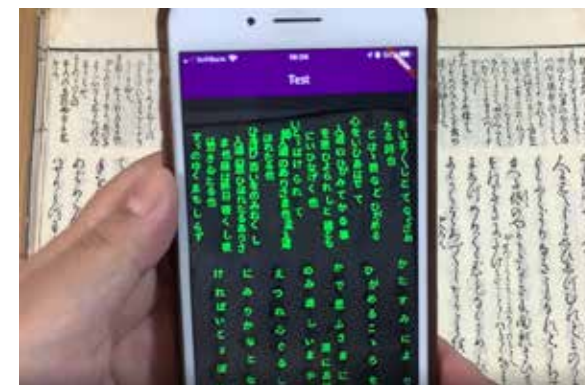
CODH カラーヌワット・タリンほか



くずし字認識サービス

kaggle

くずし字認識コンペ



くずし字認識モバイル
アプリ



NIJL-NWプロジェクト

@国文学研究資料館

<http://www.nijl.ac.jp/pages/cijproject/>

**300,000点の日本古典籍
(1868年以前) を、デジ
タル化し、オープンデー
タとして公開。**

日本文化のビッグ
データをどのよう
に活用するか？

過去の文化遺産と「くずし字」の問題

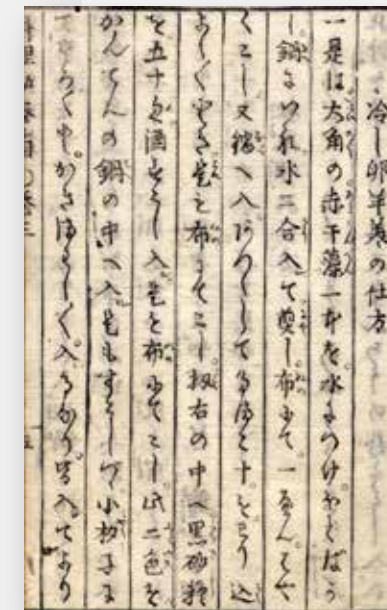


文書
数億点

日本に残された古典
籍や古文書の点数
(冊数) の推定

読者
数千人

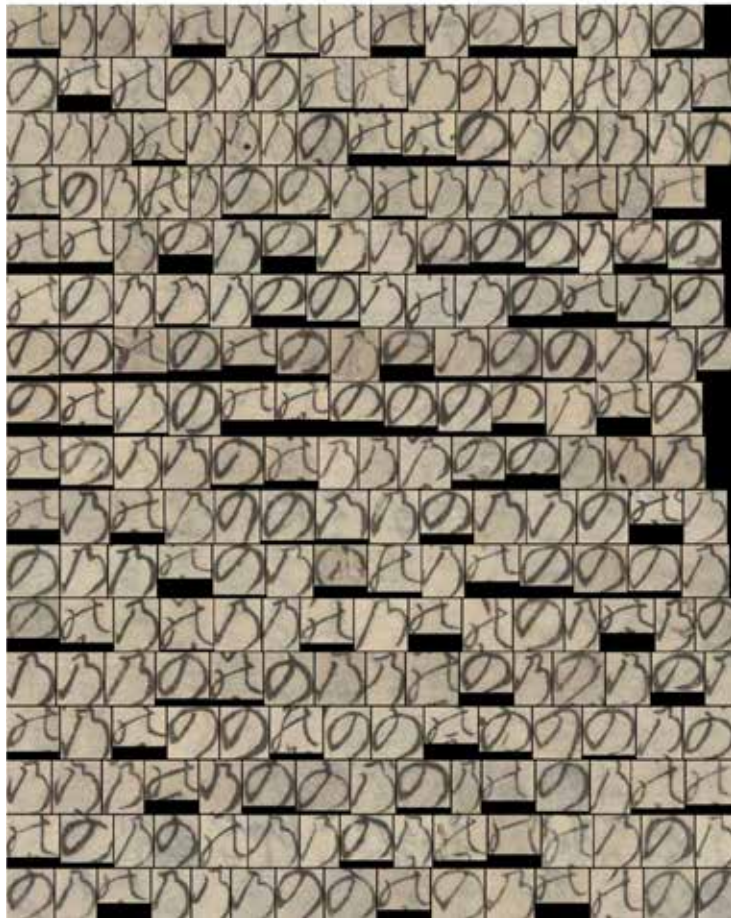
くずし字をきちんと
読める人数の推定
(**全人口の0.01%**)



くずし字データセット

<http://codh.rois.ac.jp/char-shape/>

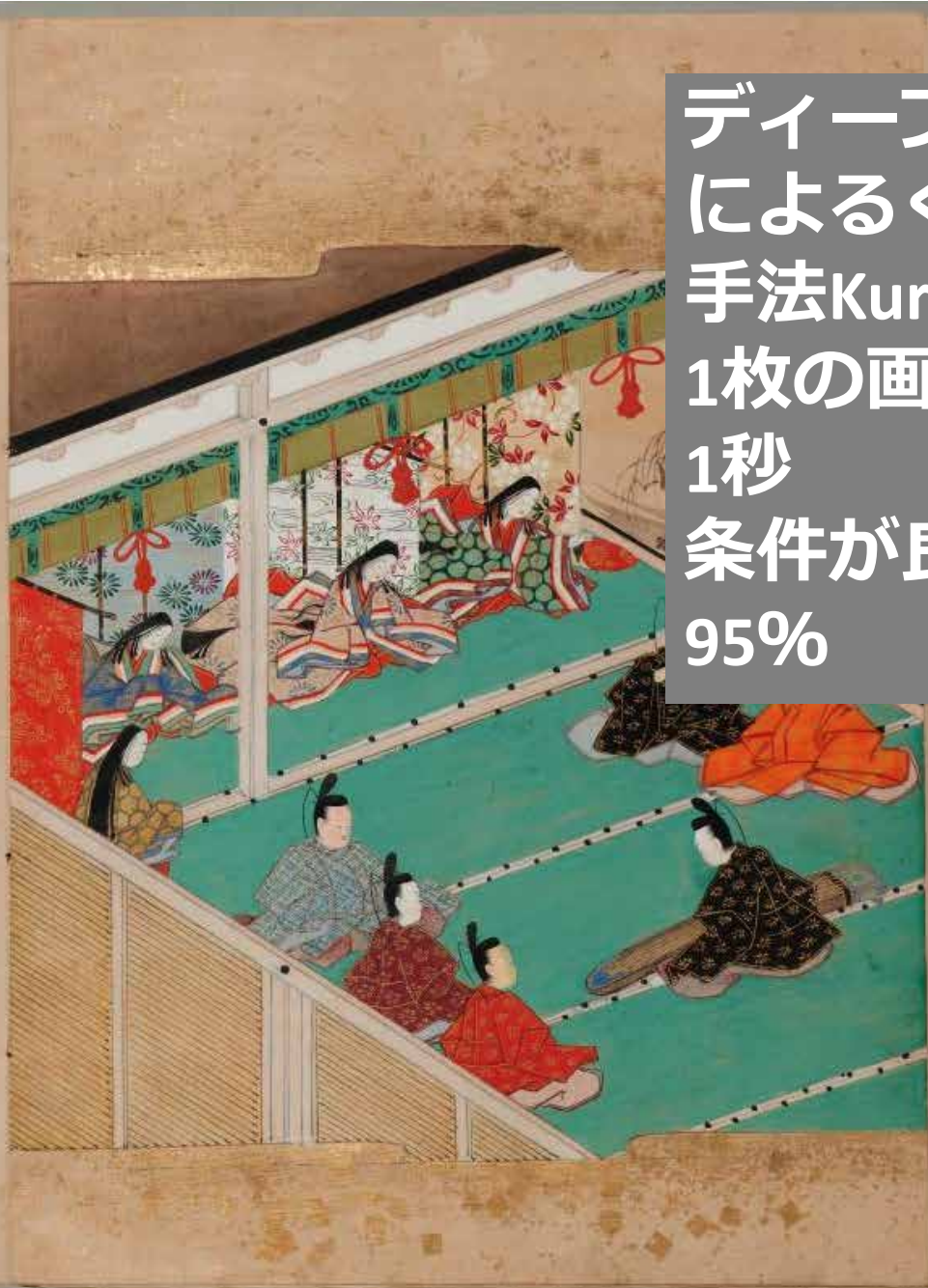
雨月物語 (1890)



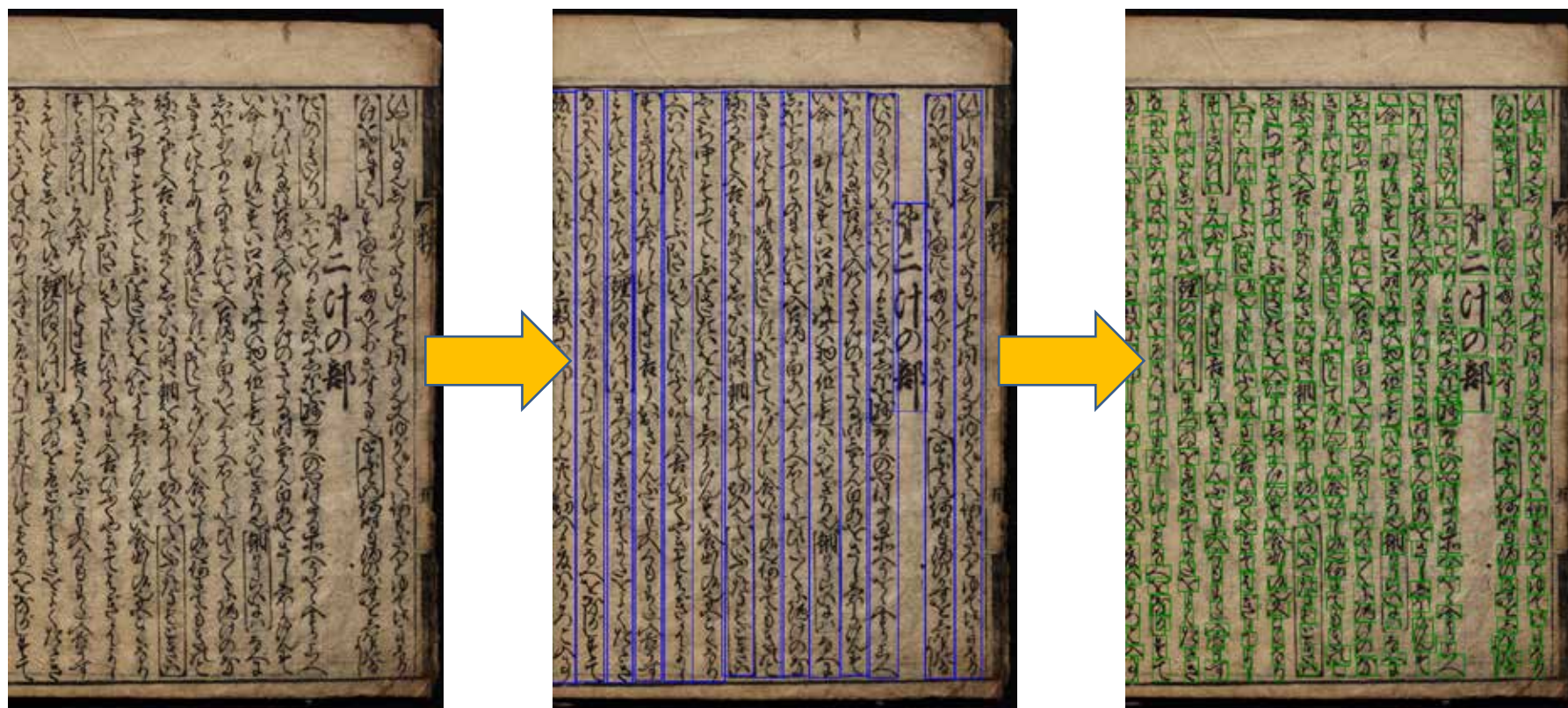
- 国文学研究資料館が作成、CODHが整理して公開するオープンデータ。
- 2020年7月現在
 - 文字種 = 4,328
 - 文字数 = 1,086,326
- ZIPファイルダウンロード→くずし字認識の訓練データとして利用。
- データセットの公開が、AIによるくずし字認識研究を活性化した。

ディープラーニング
によるくずし字認識
手法KuroNetを開発
1枚の画像の認識に約
1秒
条件が良ければ精度
95%

三てう殿に殿きたのかたならひておはし
ます御たいまいれりしううちよりまう
てたまへりくにくのしやうよりたうき
ぬぬのなともてまいれり御いそきのれう
にとてあやうす物かとりきぬなとお
ほく奉れたれはみくしけのする人御
まへまてはからひきたむそめくさ何くれの
としやうこのものともは一てう殿にもわかち
奉り給おはする事はなければ御かた
におほしなけきまくにきおとろかし



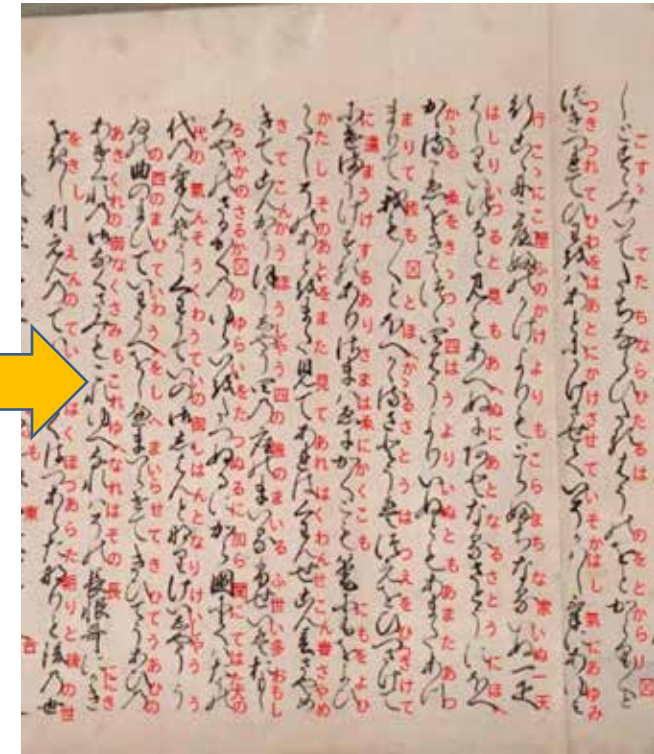
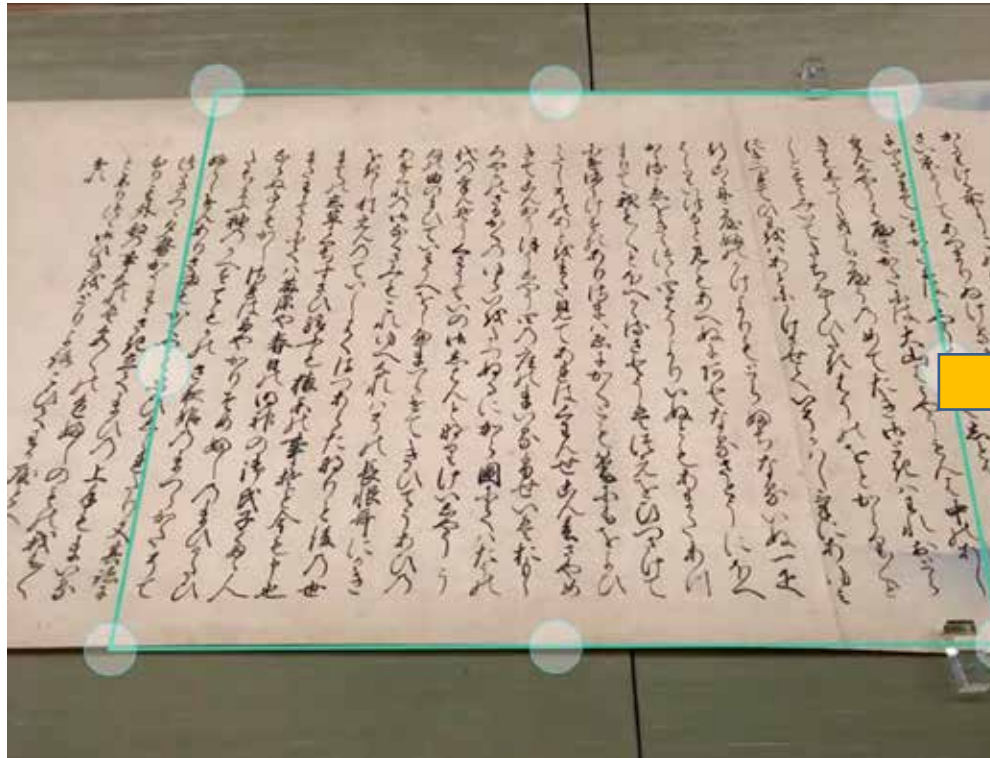
伝統的なOCR手順



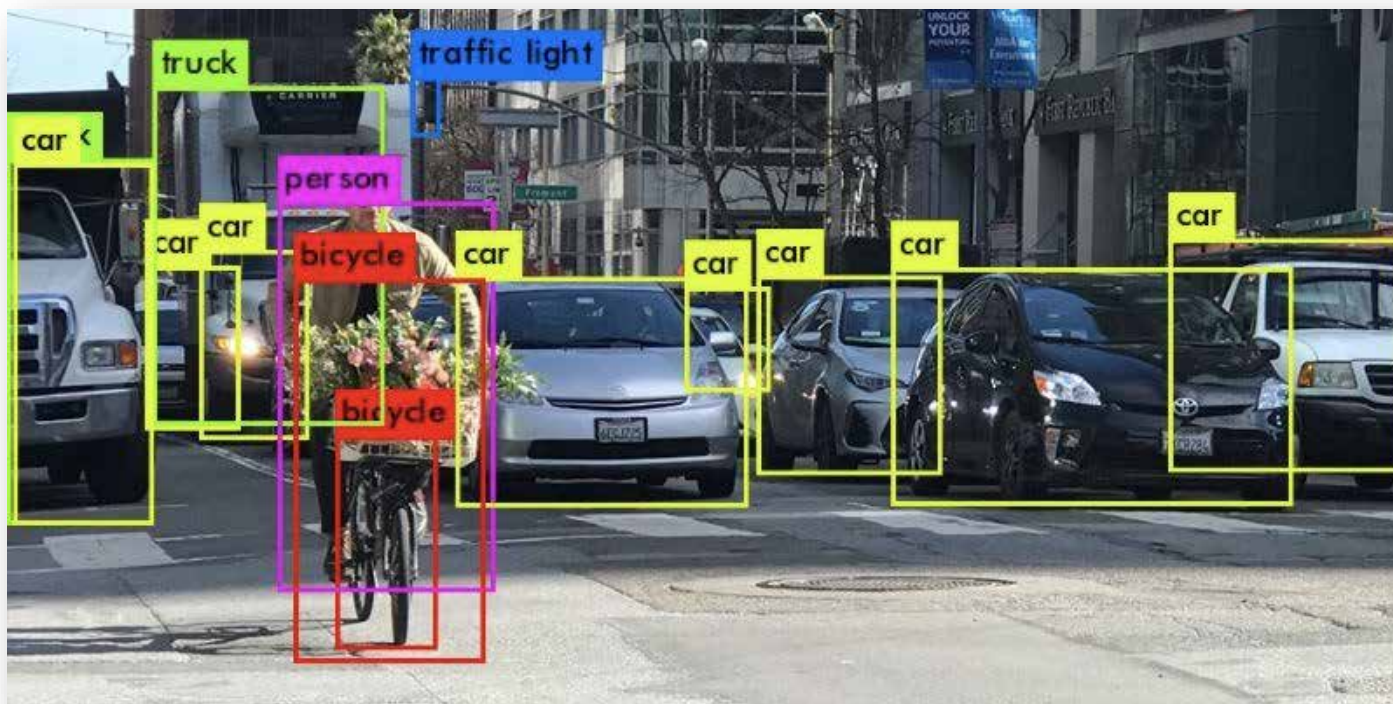
KuroNetのOCR手順（物体検出）

画像

文字認識



物体検出 (Object Detection)



1. 画像中に存在する物体を認識する技術は、自動運転などビジネスへの応用範囲も広い。
2. くずし字 = オブジェクトとみなせば、物体検出技術が使えるのでは？

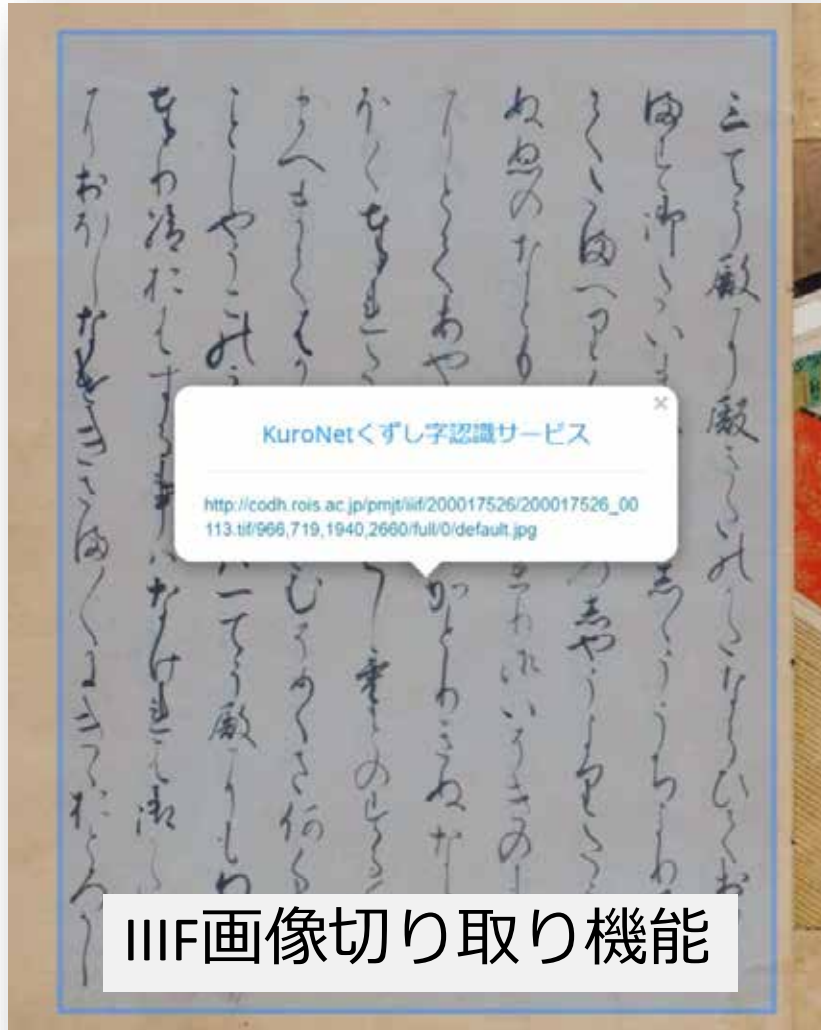
KuroNetのキーポイント

- 通常のOCR：レイアウト解析→文字認識
- KuroNet：文字認識→レイアウト解析
- 処理順を逆転させることで、レイアウト解析に依存せずに文字認識可能。
- 文字さえ認識すれば、「人間によるレイアウト解析」によりテキストが読める。
- 木版印刷はレイアウトの自由度が高いため、事前にモデル化することが困難。
- Semantic Segmentation→Object Detection

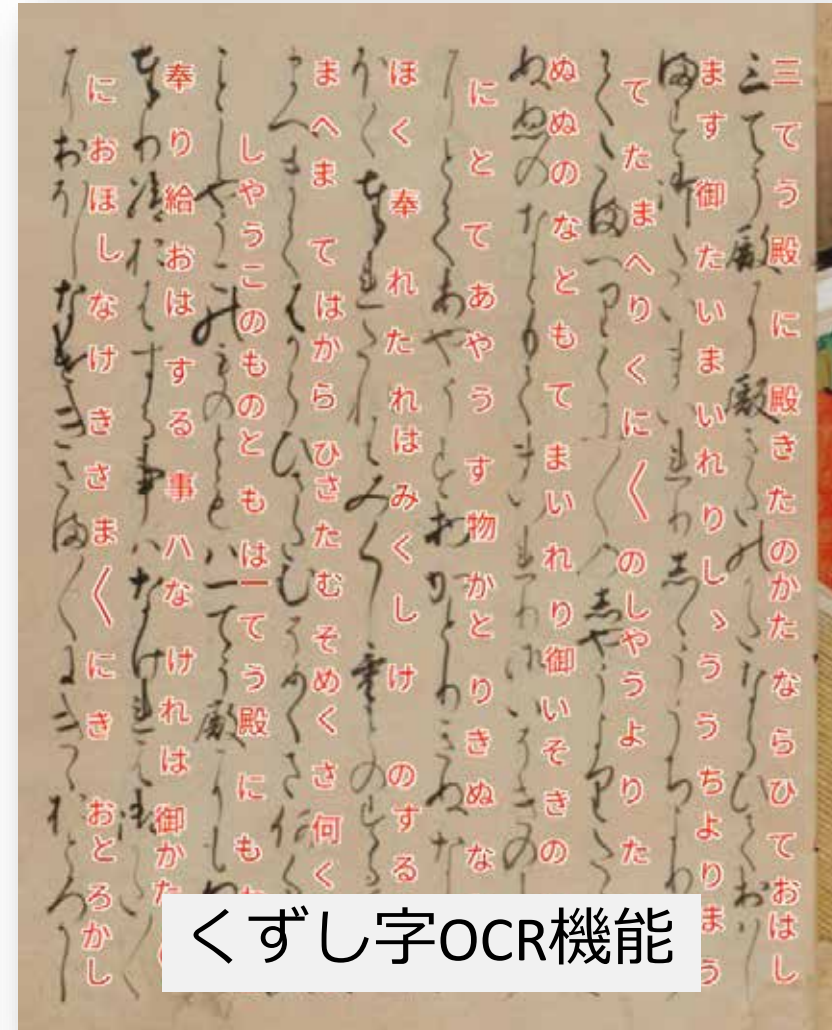


KuroNetくずし字認識サービス

<http://codh.rois.ac.jp/kuronet/>



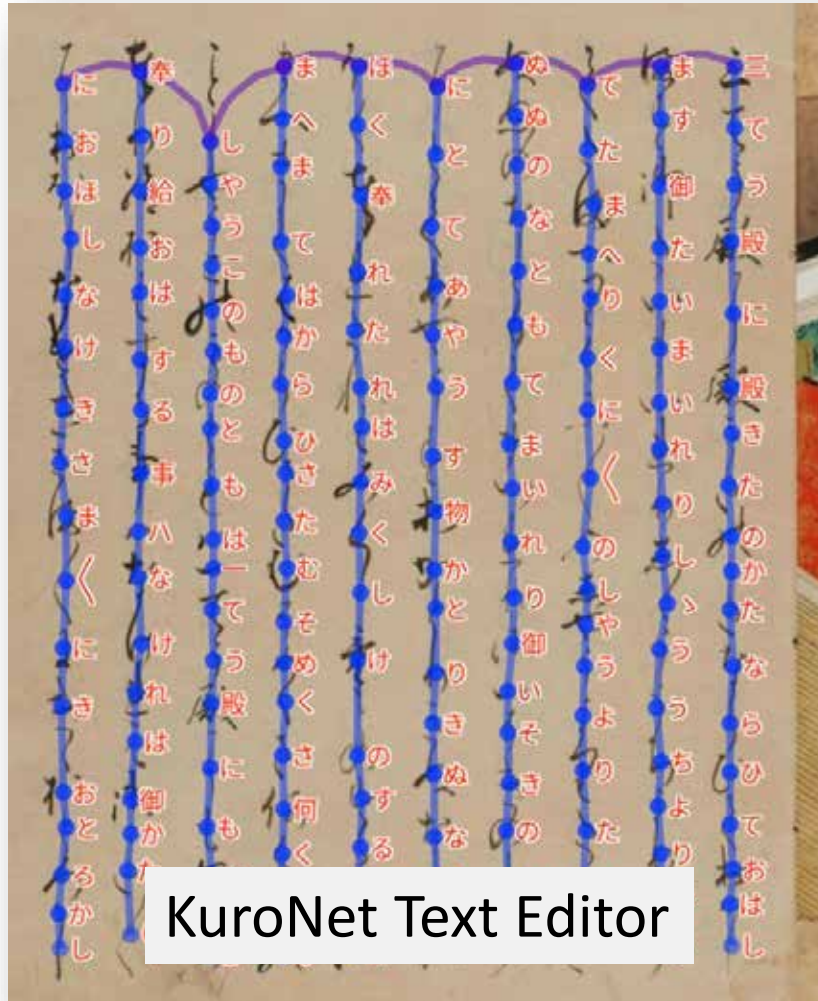
IIIF画像切り取り機能



くずし字OCR機能

KuroNetくずし字認識サービス

<http://codh.rois.ac.jp/kuronet/>



テキスト化

三てう殿に殿きたのかたならひておはし
ます御たいまいれりし、うちよりまう
てたまへりくにくのしやうよりたうき
ぬぬのなともてまいれり御いそきのれう
にとてあやうす物かとりきぬなとお
ほく奉れたれはみくしけのする人御
まへまてはからひさたむそめくさ何くれの
しやうこのものともは一てう殿にもわかち
奉り給おはする事ハなければ御かたく
におほしなけきさまくにきおとろかし

テキスト化サービス

kaggle くずし字認識コンペ

<http://codh.rois.ac.jp/competition/kaggle/>



機械学習エンジニアが300万人以上登録する、**世界最大のデータサイエンスプラットフォーム Kaggle**にて、くずし字認識コンペを開催。

- **期間**：2019年7月19日～10月14日
- **参加チーム数**：293
- **参加者数**：338
- **結果提出回数**：2652

kaggle コンペの結果

上位の精度
は約95%

1. **さまざまな国**の機械学習エンジニア、研究者が上位入賞。
2. **日本語やくずし字**の知識がなくても開発可能。
3. コンペの準備を滞りなく進めるには、**情報学者と人文学者の協働**が不可欠。

#	Δpub	Team Name	Notebook	Team members	Score 🏆	Entries
1	—	tascj			0.950	13
2	—	Konstantin Lopuhin			0.950	60
3	—	Kenji			0.944	161
4	▲1	YoudaoOCR			0.942	49
5	▼1	See--			0.940	42
6	—	abc			0.939	15
7	—	K_mat			0.934	20
8	—	t-hanya			0.920	21
9	—	Ollie, Nanashi, and Tom			0.910	35
10	—	Zenkei_R&D			0.903	144
11	—	masayai			0.903	12
12	▲5	Kirill Brodt (shad nsk)			0.901	4
13	▲1	James Day			0.901	33
14	▼1	NEU			0.900	54
15	▼3	s tatsuya			0.900	29



「みを」くずし字認識スマホアプリ

<http://codh.rois.ac.jp/miwo/>

- 『源氏物語』第14巻「みをつくし」（漣標）。
- 「みを（船の水路）を示すために立ててある杭」の意。
- 「身を尽くし」の掛け言葉。
- 「みをつくし」が人々の水先案内となるように、「みを」アプリがくずし字資料を読むための道案内となることを目指している。
- 開発者：CODHカラーヌワット・タリン氏

「みを」アプリ体験版（Kemco） 2021年4月

<http://codh.rois.ac.jp/miwo/>



日本美術

共同研究者：鈴木 親彦（CODH）、高岸 輝（東大）、Alexis Mermet（NII
インターンシップ）

顔コレデータセットの利活用例

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

日本古典籍データセット（国文研蔵ほか）



IIIF Curation Viewer / IIIF Curation Platform
(CODH開発)



顔貌コレクション・データセット（CODH作成）



美術史研究



AI自動顔認識



機械の創造性
(GAN)

CODH 鈴木 親彦、東大 高岸 輝、Google Yingtao Tian、EPFL Alexis Mermet(ほか)

オープン化のトレンド

1. **オープンデータ**：誰でも自由に使えるようデータを公開。
2. **オープンソース**：誰でも自由に使えるようソフトウェアを公開。
3. **相互運用技術**：異なるシステムで同じ技術を使い、アクセス方法を共通化。
4. **IIIF (International Image Interoperability Framework)**：3つのトレンドの交差点。

IIIF（トリプルアイエフ）とは？

International Image
Interoperability
Framework = 国際的な
画像配信方式



Web : HTML
画像 : IIIF

IIIFサー
ビス1

IIIFサー
ビス2

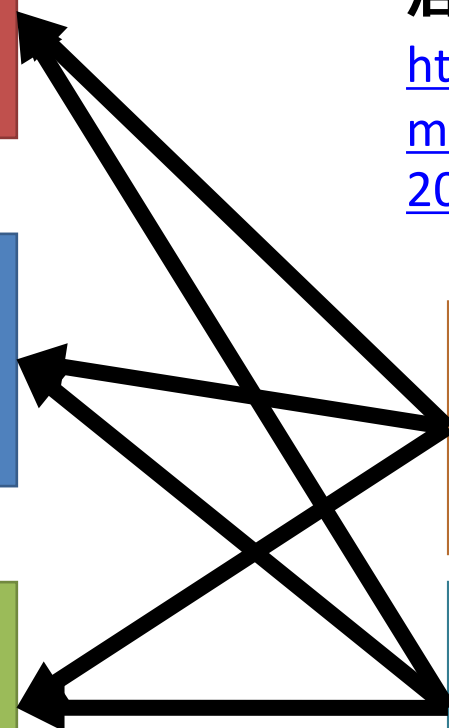
IIIFサー
ビス3

第14回CODHセミナー -
IIIF Curation Platform利
活用レシピ100連発

<http://codh.rois.ac.jp/seminar/icp-recipe-20210218/>

IIIFビュー
ア1

IIIFビュー
ア2



IIIF Curation Platform (ICP)

CODH開発

<http://codh.rois.ac.jp/icp/>



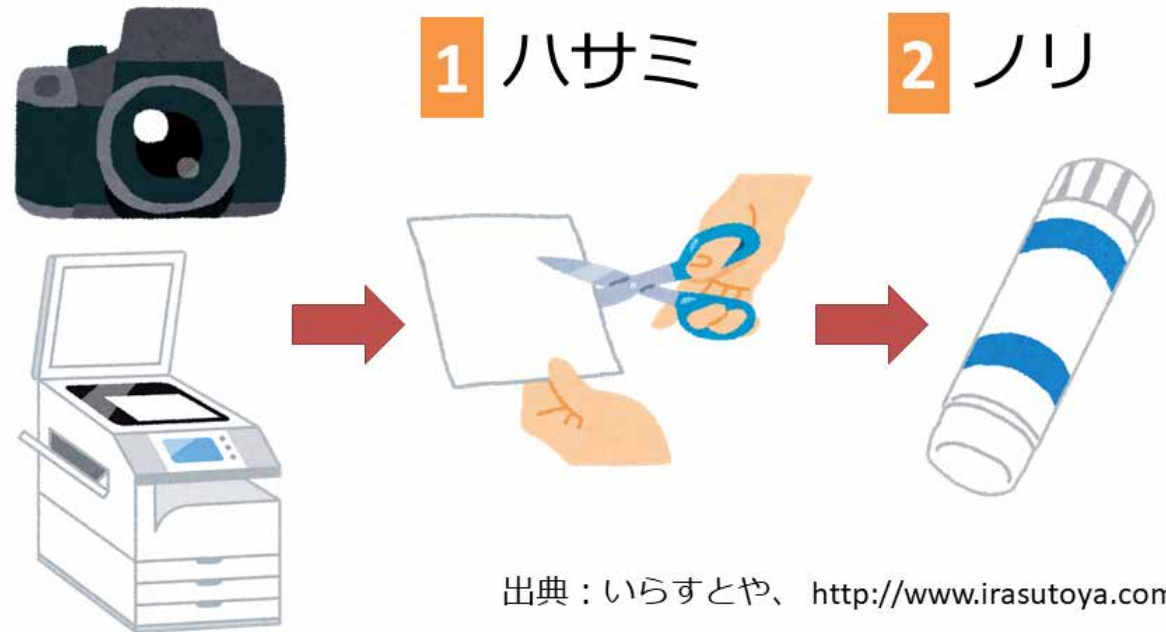
1. '■'の「切り取り」ボタン → 四角領域を指定
2. '☆'の「お気に入り」ボタン → 欲しい画像を収集



キュレーションとは？

キュレーションとは、もともと、ミュージアムにおいて、資料の収集や作品の展示などの活動を指すことばである。

1. あるテーマに沿って**コンテンツ**を集める。
2. 適切な**順番（配置）**に並べる。
3. **新たなコンテンツ**として提示・共有する。



出典：いらすとや、<http://www.irasutoya.com/>

キュレーションの作成と公開



不美人の描き方ー「未摘花」

<http://codh.rois.ac.jp/pmjt/curation/2/>



顔貌データ：イギリスの肖像

<http://codh.rois.ac.jp/curation/exhibition/2/>

顔貌コレクション (顔コレ)

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



顔コレ「山伏」の検索結果

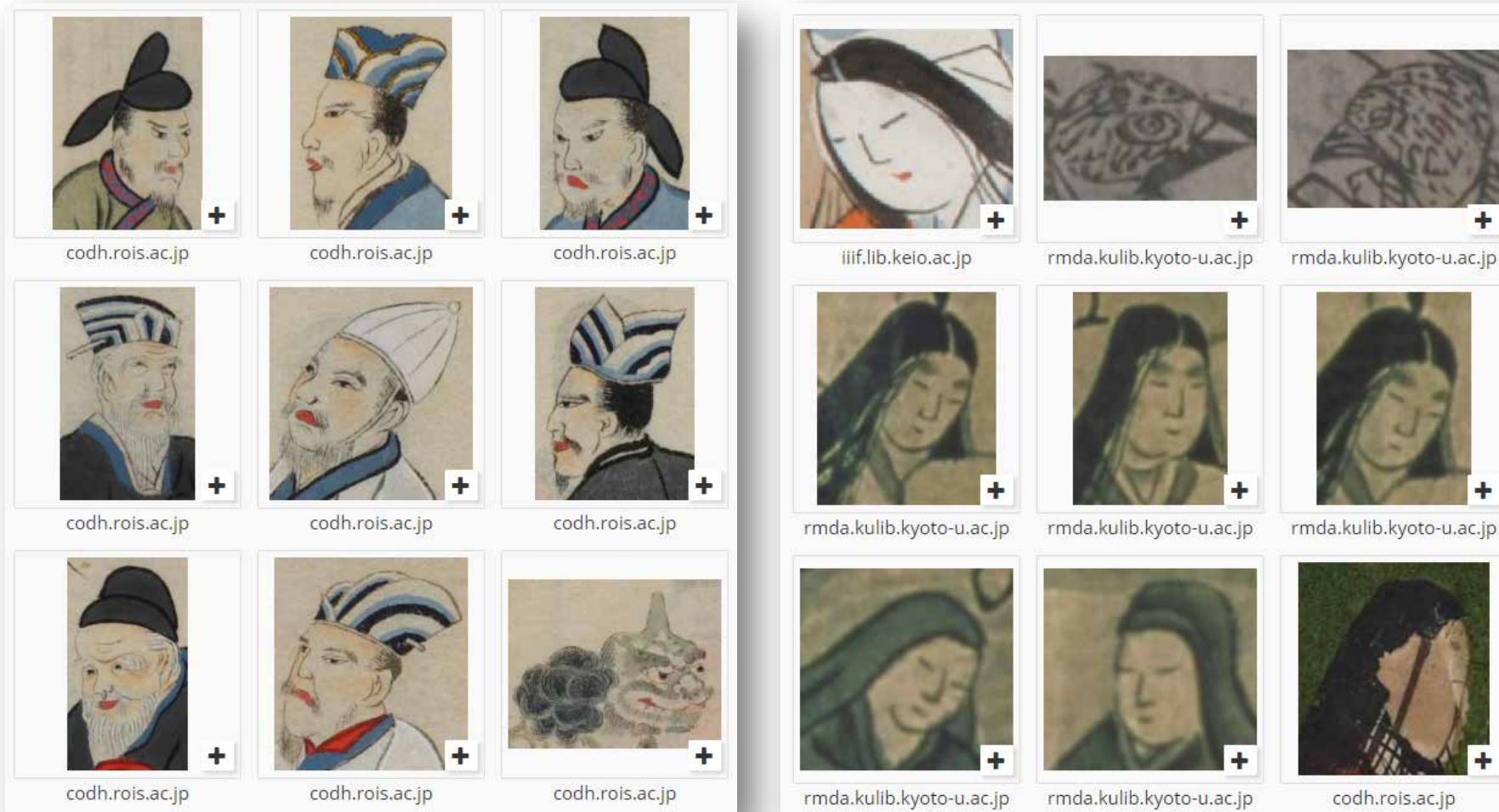
1. デジタル画像を切り抜き、メタデータを付与し、**網羅的なデータセット**を構築する。
2. **デジタルツール**を用いてデータセットを分析・可視化し、新たな発見につなげる。
3. 単にデジタルデータを使うだけではない。**研究手法もデジタル化し、その利点を活用する方向に進化させる。**

顔コレによる顔貌比較

男

<http://codh.rois.ac.jp/software/iiif-curation-finder/>

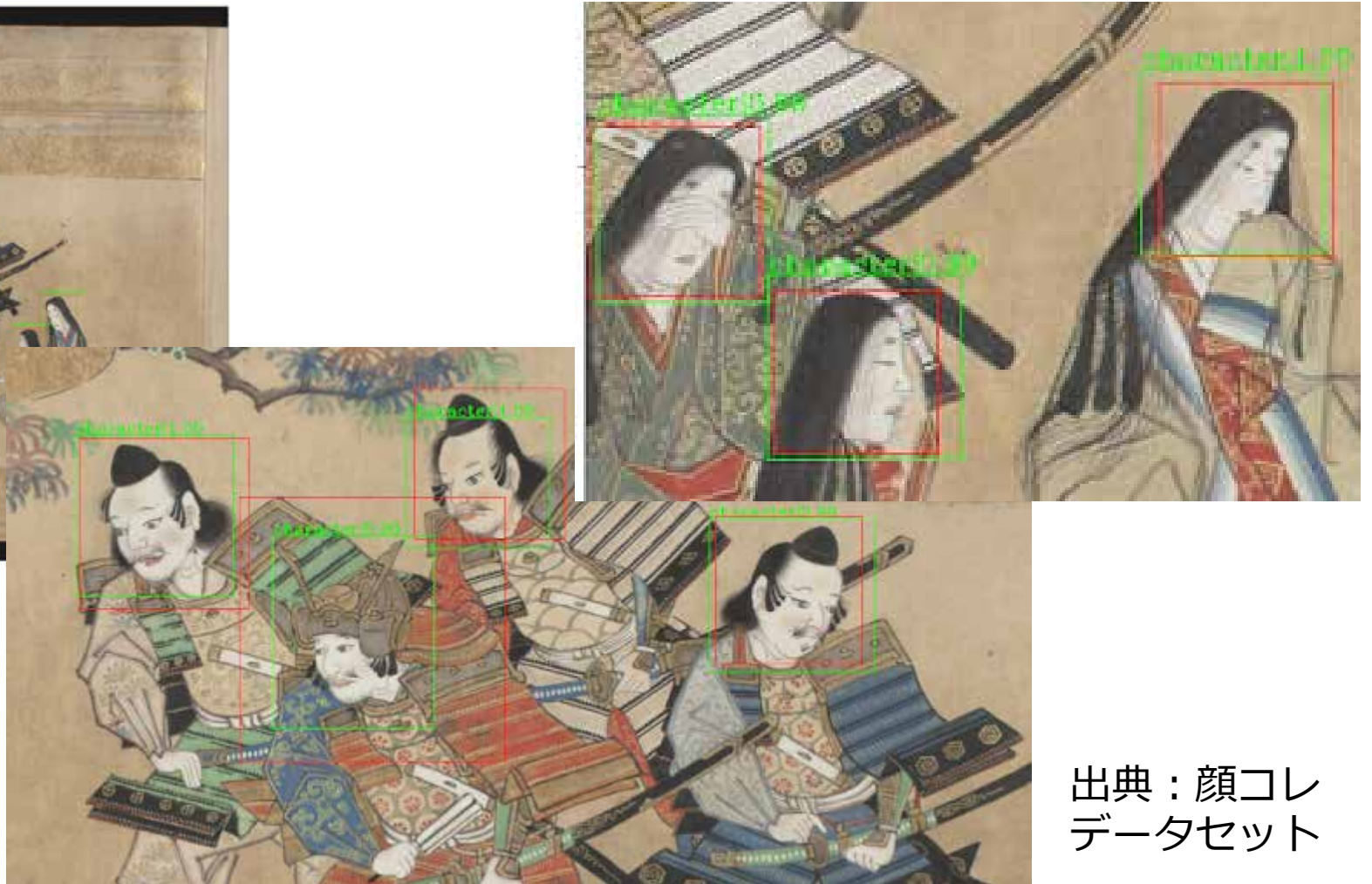
女



Faster R-CNNによる顔検出結果



Alexis Mermet, Asanobu KITAMOTO, Chikahiko SUZUKI, Akira TAKAGISHI, "Face Detection on Pre-modern Japanese Artworks using R-CNN and Image Patching for Semi-Automatic Annotation", Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritage Contents (SUMAC'20), pp. 23-31, doi:10.1145/3423323.3423412, 2020.



出典：顔コレ
データセット

機械学習による顔検出の意義

1. 顔コレデータセットを対象とした場合、**80%程度の顔を自動検出**できた。
2. 顔コレデータセットで学習し、時代が異なる作品に適用した場合でも、**70%程度の顔を自動検出**できた。
3. 3分の2の検出を機械に任せられれば、**人間が顔を切り取る作業量を3分の1に短縮**できる。
4. **より多くの作品を分析できれば、エビデンスの信頼性が向上し、研究を深めることができる。**

機械学習の役割

1. 顔の切り抜き「作業」をAIが支援することで、**作業の労力を大幅に軽減**できる可能性がある。
2. データの解釈は専門家が行う「精読」に基づくが、従来に比べて**規模を大幅に拡大**できる。
3. データの解釈も機械が支援する「遠読」については、**新たな可能性と方法論の危うさ**が共存している。
4. AIの結果は学習データに依存。「お告げ」のように**盲信し、自分に都合のよいエビデンス**としてはならない。

歴史ビッグデータ

共同研究者：鈴木 親彦、市野 美夏（CODH）、藤實 久美子（国文学研究資料館）、Thomas Leyh（NIIインターンシップ）

歴史ビッグデータの統合解析

<http://codh.rois.ac.jp/historical-big-data/>

過去のビッグデータを統合解析するための基盤技術进行研究



自然科学的
データ

人文
社会的
データ

気候

地震

噴火

疫病

経済

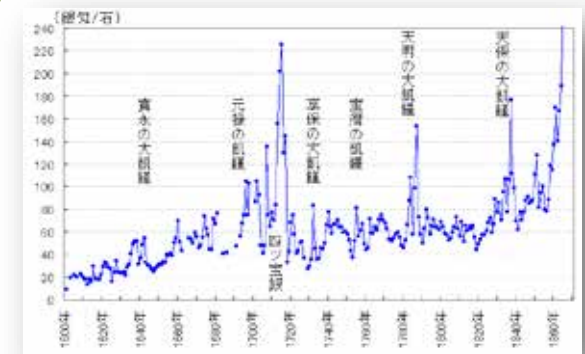
人口

政治

文化

データ
構造化
ワーク
フロー

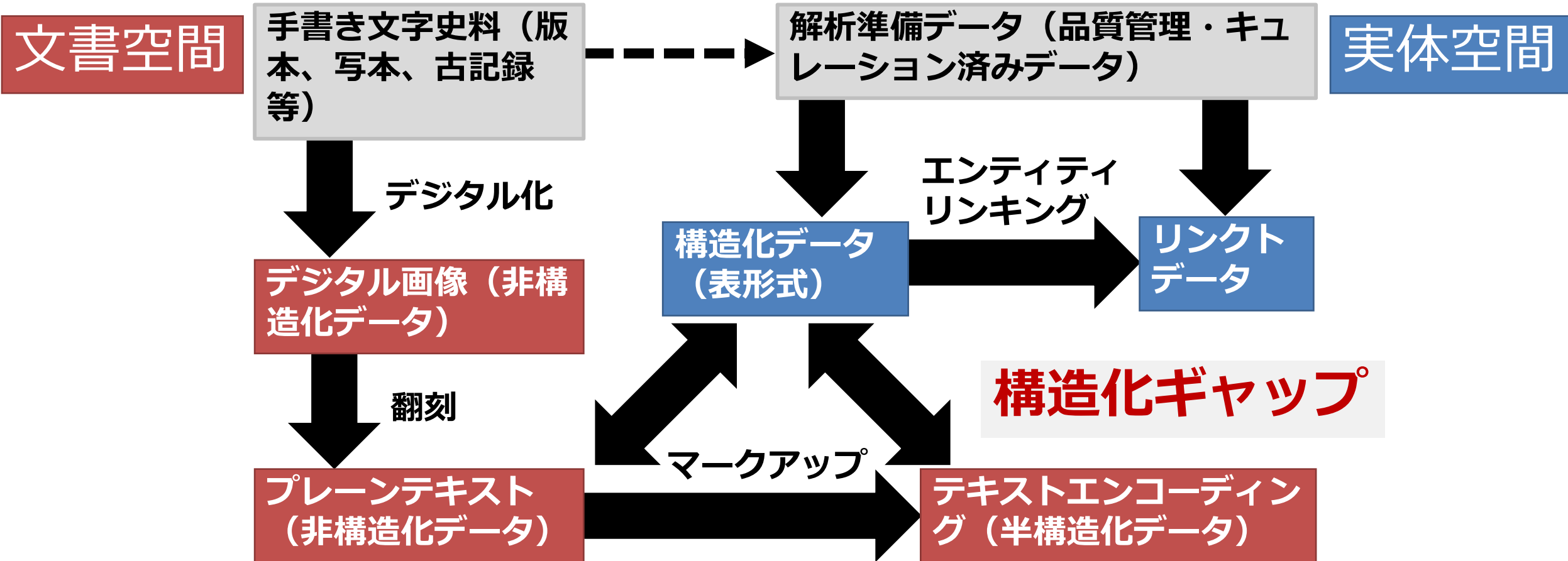
歴史ビッグ
データ研究基
盤（機械可読
データ）



歴史ビッグデータとは？

1. 現代のデータを想定して開発されたビッグデータの方法論を、過去の記録に延長すること。
2. 生データの構造化：機械可読なデータを生成し、様々な分野における解析などに利用する。
3. 多様な（Variety）データの統合解析：単体のデータだけでは見えてこない新しい知見を得る。
4. 人文学研究のデジタルトランスフォーメーション：研究のデジタル化には、技術要素の深化だけでなく、研究体制や（評価）制度の変革も関わる。

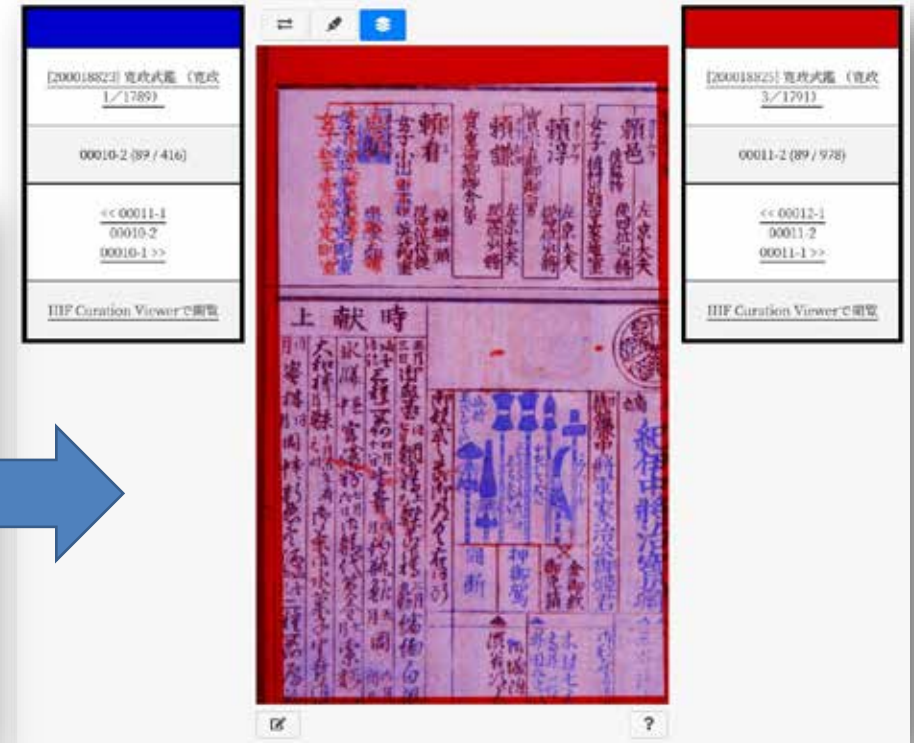
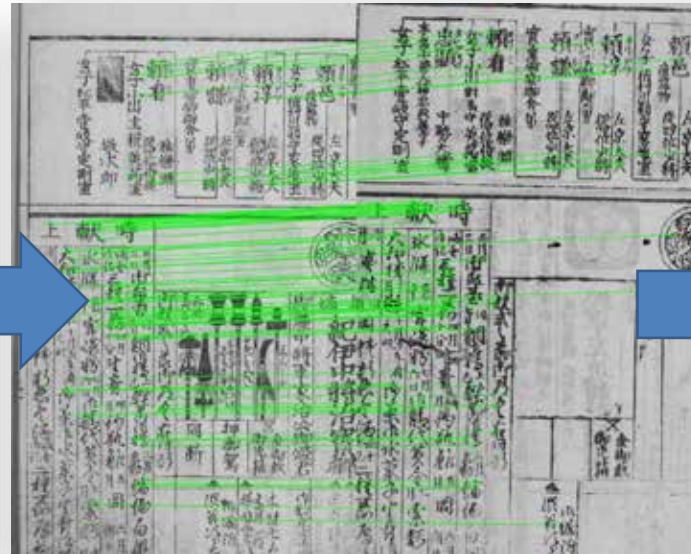
データ構造化ワークフロー



武鑑全集の構造化

<http://codh.rois.ac.jp/bukan/>

多数の版を視覚的に比較して差分を取り出す



江戸時代に頻繁に更新され出版された、大名・幕府役人に関するデータブック（国文研蔵）

コンピュータビジョン技術を用いた2枚の画像のマッチング

版間の差分を強調した「差読」の実現

「武鑑」とは？



寛政武鑑（1789）、日本古典籍
データセット（国文研所蔵）

<http://codh.rois.ac.jp/pmjt/book/200018823/>

1. 「武鑑」とは、江戸時代に出版された大名家および幕府役人の名鑑。
2. 17世紀中ごろ～1867年の大政奉還まで、200年以上の間出版され続けた。
3. 実用書であり、ロングセラーブック。
4. 社会の需要に応じて、年を追うごとに厚くなり、その改訂の頻度は年に数度から月に数度までに増えた。

参考：藤實 久美子：江戸の武家名鑑 武鑑と出版競争, 吉川弘文館, 2008

武鑑全集

<http://codh.rois.ac.jp/bukan/>

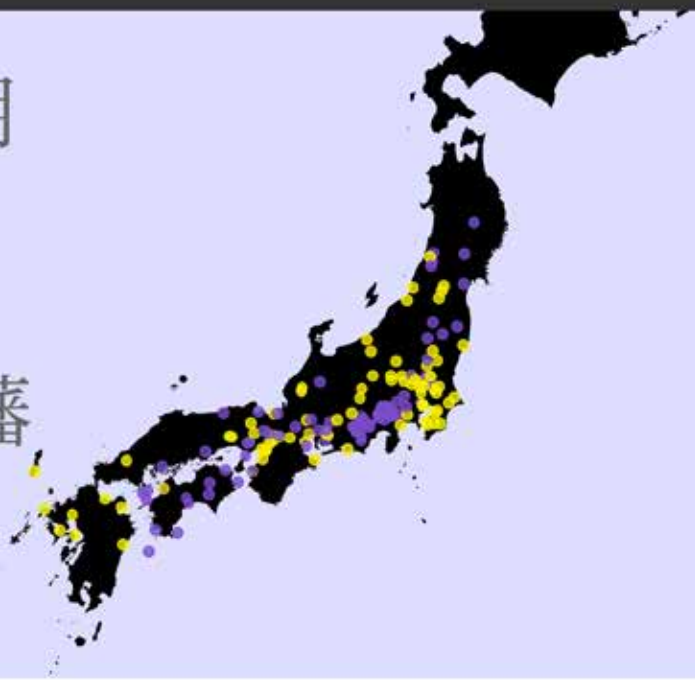
武鑑全集



Edo+150

参勤交代アニメーション

干支：子4月
参府：42藩
暇：43藩
江戸：129藩



武鑑全集



Edo+150



大名家一覧

日本古典籍データセットで公開する寛政武鑑（1789）の大名家一覧です。IDは寛政武鑑（1789）での出現順に付与しています。

ID	大名当主名（現代通称）	大名当主名（武鑑表記）	藩名（現代通称）	居城地（武鑑表記）	領知高（単位：石）	参勤交代年月（参府）	参勤交代年月（暇）
1	徳川宗睦	尾張大納言宗睦	尾張	尾州愛知郡名古屋	619,500	子寅辰午申戌 3月	丑卯巳未酉亥 3月
2	松平義祐	松平摂津守義祐	高須	澁州石津郡高須	30,000	子寅辰午申戌 4月	丑卯巳未酉亥 4月
3	徳川治貞	紀伊中納言治貞	紀州	紀州名草郡和歌山	555,000	丑卯巳未酉亥 3月	子寅辰午申戌



大名家一覧

参勤交代マップ

木版印刷の版本とバージョンの関係



左『寛政武鑑』 (1789)



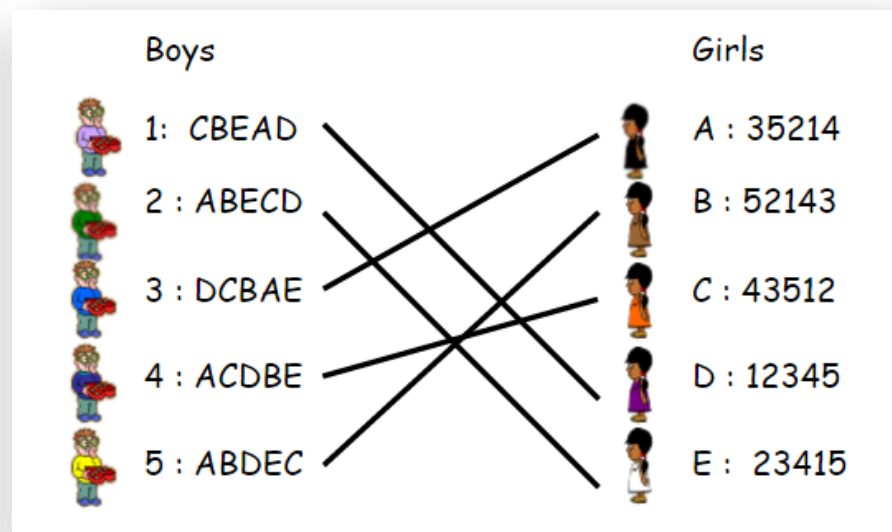
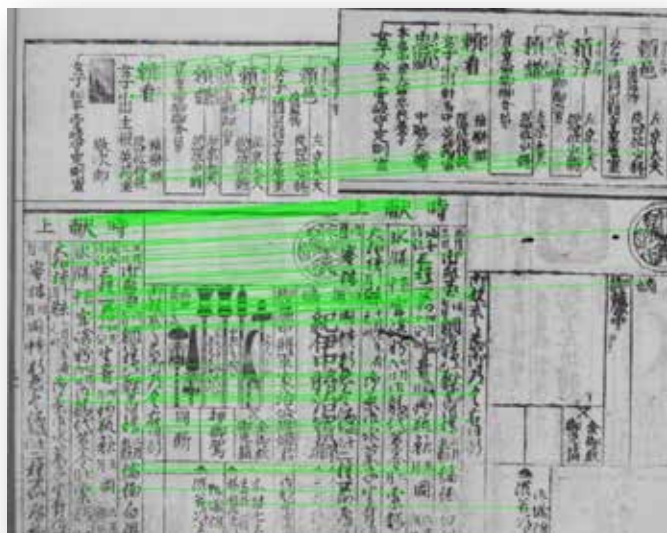
右『寛政武鑑』 (1791)

江戸時代の出版は、板木に文字や絵を彫って印刷する**木版印刷**が主流。

1. **刊（板・版）**：新しい板木を彫って本を刊行すること（**メジャーバージョン**）。
2. **印（刷・摺）**：既存の板木を使って本を刷ること（異なるバイナリ？）。
3. **修（補・訂）**：既存の板木に対して埋木などを使って部分的な修正を加えること（**マイナーバージョン**）。

参考：中野 三敏：書誌学談義 江戸の板本, 岩波書店, 2015

ブックバーコーディング法



ページ間のマッチング：画像特徴点の抽出とマッチング（AKAZE等）

版本間のマッチング：安定結婚問題（Gale-Shapleyアルゴリズム）

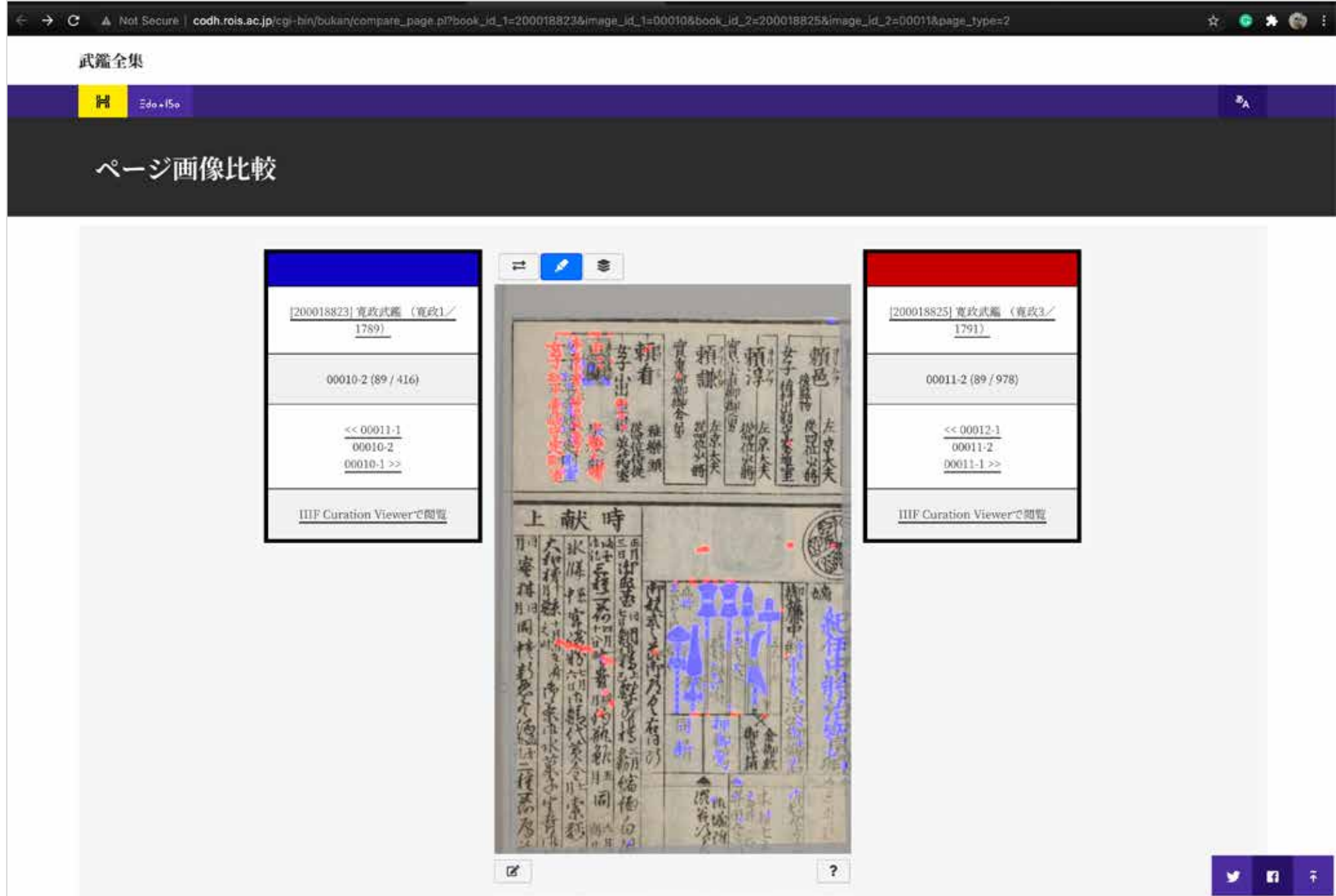
Figure: Bjarke Ebert,
<https://towardsdatascience.com/stable-matching-as-a-game-a68c279d70b>

ユニークな遺伝子配列により生物種を同定する「DNAバーコーディング法」のアナロジー

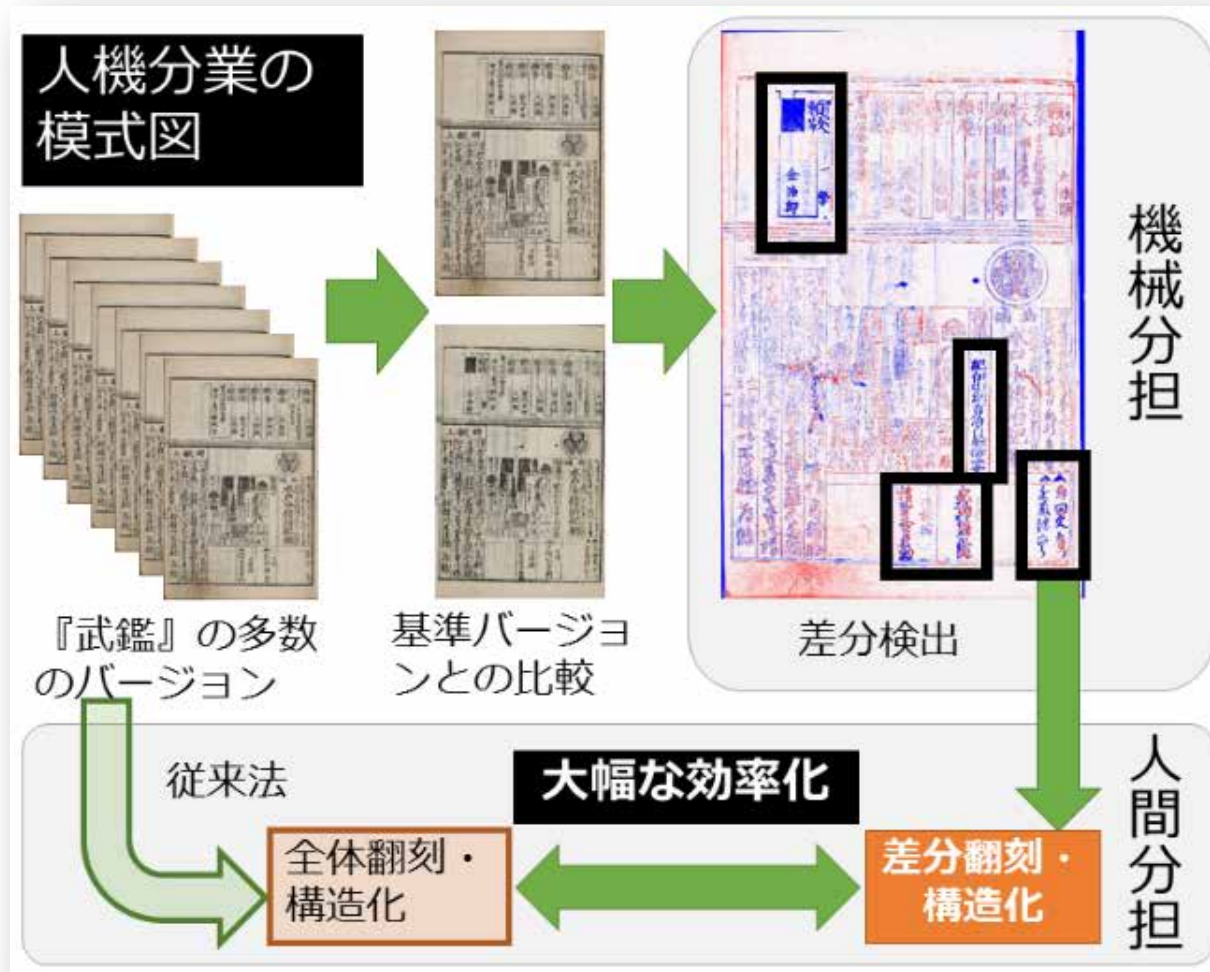
1. ページごとのユニークな特徴点の配置を用いて、画像や版本をマッチング
2. 版本ごとのユニークなコードをデータベース化し、未知の版本の「系統樹」上の位置を同定

画像比較 ツール vdiff.js

<http://codh.rois.ac.jp/software/vdiff.js/>



差読と人機分業ワークフロー



1. **差読（differential reading）**：画像の「間違い探し」という苦痛な作業を、**コンピュータの助けによって簡単な作業に反転させる**。
2. **差分翻刻**：版間で**変化した部分だけを翻刻**すれば、作業量が大幅に減る。
3. **人機分業**：**苦痛な単純作業**は機械が行い、**知識が必要な高度作業**を人間が行う。

江戸ビッグデータの統合

<http://codh.rois.ac.jp/edo+150/>

画像：江戸切絵図（国立国会図書館）

地名：CODHがデータクリエイターと共に、
IIF画像への注釈としてデータ整備



CODH 鈴木 親彦ほか

地名識別子の整備・共有 (CODH)



GeoLOD

<https://geolod.ex.nii.ac.jp/>



現代地図との重ね合わせ



商業ビッグデータ



観光ビッグデータ

江戸マップ－江戸切絵図の地名リソース化

<http://codh.rois.ac.jp/edo-maps/>



番号	分類	現代語訳	翻刻	地図
2-001	施設	幸橋御門	幸橋御門	拡大図
2-002	施設	山下御門	山下御門	拡大図
2-003	施設	数寄屋橋御門	数寄屋橋御門	拡大図
2-004	施設	鍛冶橋御門	鍛冶橋御門	拡大図
2-005	施設	呉服橋御門	呉服橋御門	拡大図
2-006	地名	一石橋	一石橋	拡大図
2-007	地名	出橋	出橋	拡大図
2-008	町名	丸屋町	丸屋丁	拡大図

[2-296]
地名：磯辺大神宮（イソベ大神宮）
分類：寺社仏閣

29枚の地図から
8658ヶ所の地名を抽出して
データベース化

2022/10/13

諸科学集

ICViewer | 国立国会図書館 National Diet Library, JAPAN

古地図のジオレファレンス



国立国会図書館
『江戸切絵図』

GCPの対応付け



立命館大学
日本版Map Warper



江戸マップβ版 + 日本版Map
Warperタイル配信サービス



千代田区

© 2020 ZENRIN

科学集会

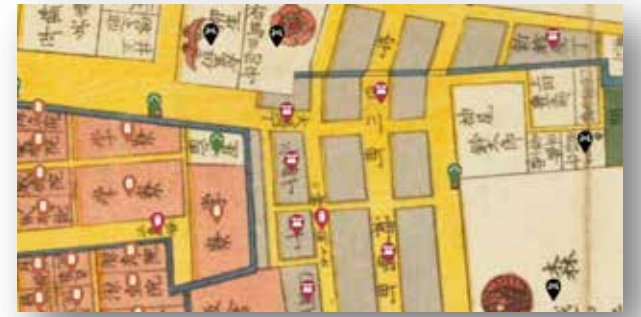
Google Earth

2021/8/18

1997

35° 40'52.26" N 139° 45'26.42" E 標高 200 m

史料とエンティティのリンク



宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

固有表現認識

宇田川丁、三島丁・
神明丁、此分潰家多、
土蔵残所なし

御江戸大地震大破并出火類焼場等書上之写（みんなで翻刻）

1. 江戸マップに出現する地名を、**地名リソース（エンティティのデータベース）**として整備する。
2. 史料の文字列から**固有表現（地名）**を抽出する。
3. **特定のエンティティとリンク**することで、実世界と紐づける。

曖昧性解消

原資料表記	江戸マップID	江戸マップ表記
宇田川丁	4-358	宇田川町
三島丁	4-290	三島丁
神明丁	4-294	神明町

江戸買物案内 – 商業ビッグデータ

<http://codh.rois.ac.jp/edo-shops/>



- 『江戸買物独案内』（1824）から、商人名や居所などを抽出しデータベース化。
- 歴博の「江戸商人・職人データベース」は現在の区レベル。
- 江戸買物案内は、江戸マップの町レベルでリンク。

江戸観光案内－観光ビッグデータ

<http://codh.rois.ac.jp/edo-spots/>

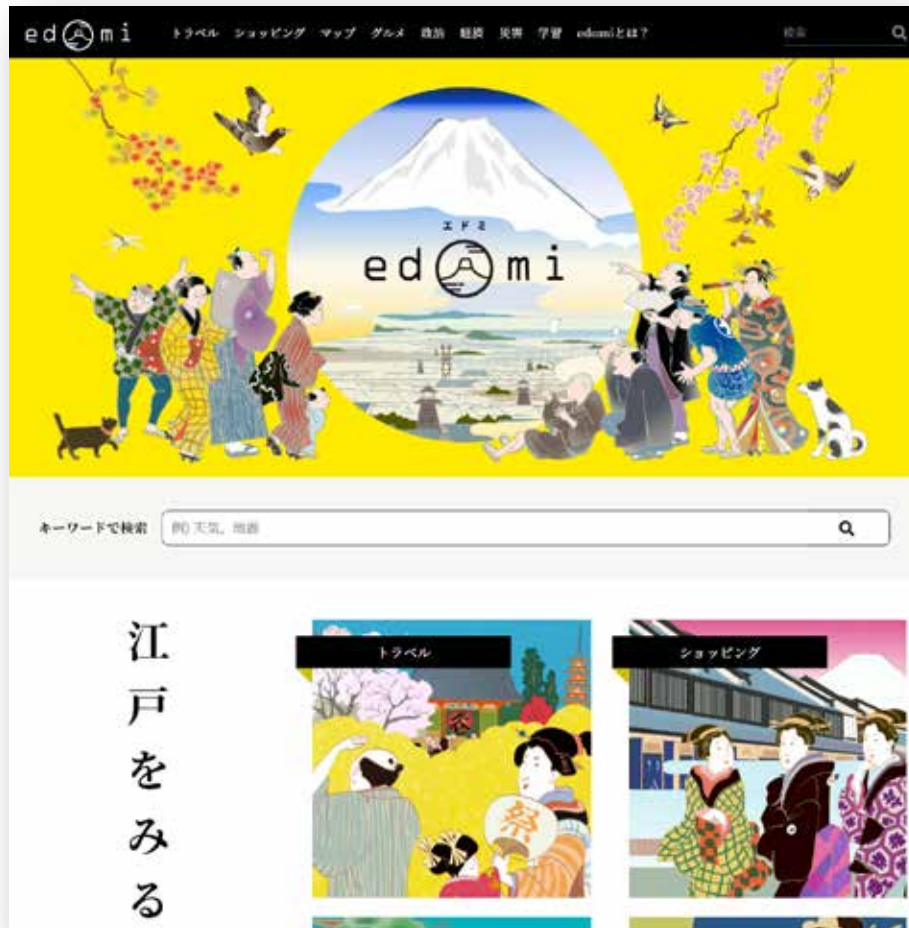


- 江戸時代の名所記や名所案内を各世紀2点ずつ選択。
- IIF Curation Viewerを用いて挿絵を切り抜き、地名を翻刻し、メタデータを付与。
- 江戸マップ、歴史地名データとリンク。

CODH 鈴木 親彦 ほか

edomi - 江戸をみる／みせるデータポータル

<http://codh.rois.ac.jp/edomi/>



1. 江戸の都市、江戸の時代に関する**構造化データを公開するデータポータル**。
2. **現代の地図、現代の分類体系**など、現代の標準と過去のデータを接続する。
3. 現代のYahoo!のように、**カテゴリごとにデータを収集し、整理し、公開する**。

おわりに

人間と機械が共に読む時代

1. 過去の膨大な文書は、人間が読むために作られてきた。
2. これからは人間と共に機械も読み、相互に助け合う時代になる。
3. 機械可読形式への変換技術基盤が必要（ビジネス分野ではRPAとも呼ばれる）。
4. AI（機械）は意味が理解できない。文化を読み解くには、人間の知識が必要。

AI（機械）は人間の代替なのか？

1. **最善主義（best effort）**：人間の苦痛を軽減するツールとしての価値は高い。
2. **完璧主義（perfection）**：例外も適切に扱って人間並みに精度を高めるのは困難。
3. **「苦痛の軽減」と「価値の増大」**：前者を最善主義で進めれば、AIが効果を発揮する可能性が高い。
4. **研究レベルの向上**：研究者の苦痛が軽減され、本来やるべきことに集中できれば、価値の増大に貢献できる。

日本文化 = 我々の資産

日本文化を資産として活用するには、情報プラットフォームの構築と専門人材の育成が必要。

1. 資料をデジタル化し、文字認識し、全文検索。
2. 機械と市民と専門家が得意な部分で協力。
3. 過去から現代への文化の連続性を一望。
4. 学術的な価値に加えて経済的な価値にも貢献。
5. 危機への備え = 過去の世界に関する社会の記録から、未来に向けた教訓を学んでいく。

謝辞

<http://codh.rois.ac.jp/>

- ROIS-DS CODHのメンバー（鈴木 親彦、市野 美夏、カラーヌワット タリン）の研究成果を紹介しています。
- 共同研究者の、山本 和明、藤實 久美子（国文学研究資料館）、高岸 輝（東京大学）、Alex Lamb（モントリオール大学）、Mikel Bober-Irizar（ケンブリッジ大学）らに感謝します。
- IIF Curation Platformの開発には、本間 淳（フェリックス・スタイル）、Tarek Saier（カールスルーエ工科大学）が貢献しています。
- NIIインターンシップ制度で来日した、Alexis Mermet（スイス連邦工科大学ローザンヌ校）、Thomas Leyh（フライブルグ大学）の研究成果を紹介しています。
- 研究費として、科研費 基盤(A)「歴史ビッグデータ研究基盤による過去世界のデータ駆動型復元と統合解析」、情報・システム研究機構 機構間連携・文理融合プロジェクト「人文学ビッグデータにおける構造化ギャップの克服と分野横断的利用の検証」などの支援を受けています。