

# 人文学資料マイクロコン テンツ化による情報学・ 人文学の共同研究



ROIS-DS人文学オープンデータ共同利用センター  
特任助教 鈴木親彦

# 本日の報告内容

## 0. CODHの紹介

1. 人文学資料マイクロコンテンツ化とは
2. マイクロコンテンツ化の実践：顔貌コレクション（顔コレ）
3. 顔コレを軸とした共同研究事例
4. 研究手法の確立とさらなる展開

# 人文学オープンデータ共同利用センター (CODH)

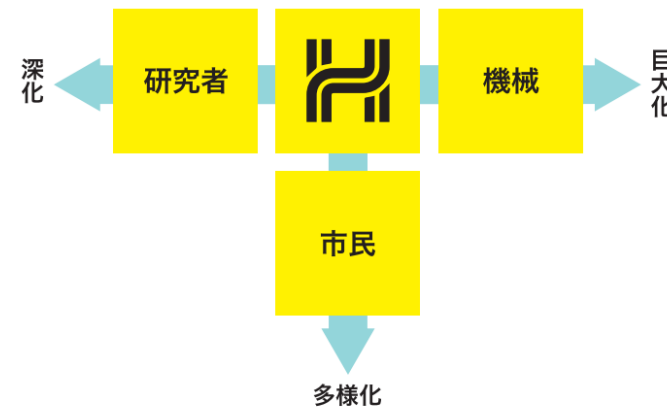
- <http://codh.rois.ac.jp/>
- 目的：人文学分野におけるデータ駆動型研究の推進と共同利用拠点の形成
- メンバー：センター長＋特任助教4名、国立情報学研究所・統計数理研究所の所属研究者



ROIS-DS人文学オープンデータ共同利用センター  
<http://codh.rois.ac.jp>

## ROIS-DS人文学オープンデータ共同利用センターの概要

“データ駆動型人文学と人文学  
ビッグデータに挑む”



「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。

ROIS-DS人文学オープンデータ共同利用センター (Center for Open Data in the Humanities / CODH) は、情報学・統計学の最新技術を用いて人文学資料 (史料) を分析する「データ駆動型人文学」や、人文学研究の成果に基づき構築したデータセットを超学際的に活用する「人文学ビッグデータ」など、オープンサイエンス時代の新しい人文学研究を展開します。

CODHではオープンサイエンスを支える三者、すなわち知を深める役割を果たす「研究者」、知を大きくする役割を果たす「機械 (AI)」、そして知を多様化する役割を果たす「市民」に注目し、それぞれに異なる強みをもつ三者が協働することで人文学を発展させるための情報学的基盤の構築を進めています。

またCODHでは、多くのオープンデータ、オープンソースソフトウェアを公開することで、人文情報学 (デジタルヒューマニティーズ) をはじめとする様々な分野の研究を加速します。さらに機械学習のコンペティションを通じて知の共有を世界的に推進する

# CODHで進める研究・二つの柱

## 人文学研究の変革

データサイエンスに基づく人文学（人文情報学）研究の促進。データサイエンスによる研究支援や、新しい人文学研究手法の確立を目指す。

その一つとして…

「人文学資料マイクロコンテンツ化」

## 人文学オープンデータの展開

人文学オープンデータを人文学以外の学問分野（情報・災害・気象…）にも利用可能にし、分野を超えた新たな研究・発見につなげる。



## 人文学オープンデータ

# 人文学資料マイクロコンテンツ化

## 「マイクロコンテンツ」

- 電子書籍から、特定の記事や特集のみを切り出して販売されるコンテンツ。興味のある部分だけを安価・手軽に読むことができる。（『デジタル大辞泉』より）

## 「人文学資料マイクロコンテンツ化」

- デジタル化された人文学資料から、特定の目的を持って部分を切り出し活用可能にすること（鈴木による造語）
  - 科学研究費助成「人文学資料マイクロコンテンツ化の実践研究－江戸の都市空間再構築を通して－」研究代表者：鈴木親彦（20K20141）

# 画像 (非文字) 資料への注目

- テキスト (文字) 情報の抽出方法としては、すでに人文情報学のデファクトスタンダードがある
  - TEI (Text Encoding Initiative) <https://tei-c.org/>
- 画像 (絵画・挿絵 etc.) を対象とした手法は研究段階
  - しかし、画像を含む資料のデジタル公開は近年盛んになっている
- 特に、相互運用性が担保されたIIIF (International Image Interoperability Framework) に則った画像公開が進展
  - 組織・公開サイトを横断して画像資料を相互運用可能に
  - <https://iiif.io/>

# 画像資料の様々（すべてIIF準拠で公開されている）



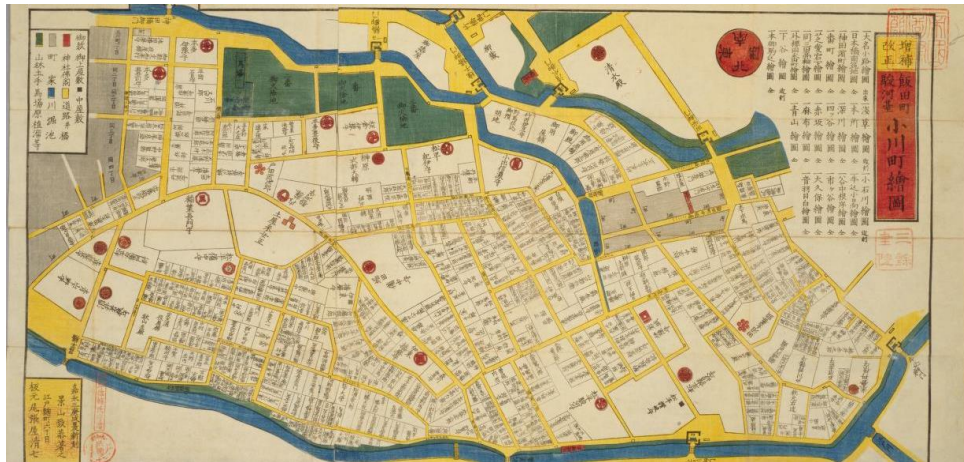
絵画『源氏物語団扇画帖』（日本古典籍データセット・国文学研究資料館蔵）



絵巻物『大黒舞』（日本古典籍データセット・国文学研究資料館蔵）



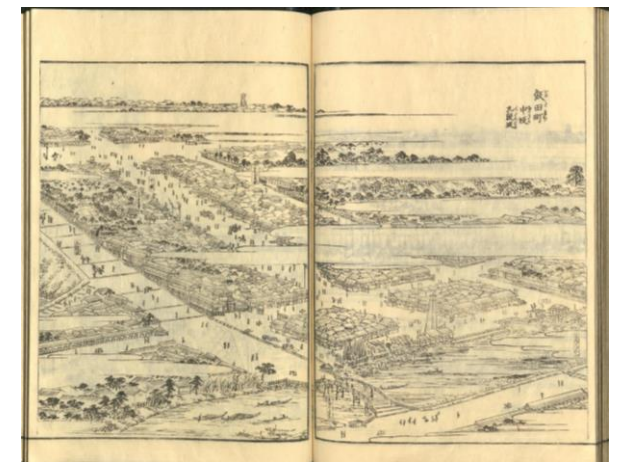
浮世絵『俳優似顔東錦絵』（国立国会図書館デジタルコレクション）



地図（絵図）「江戸切絵図」（国立国会図書館デジタルコレクション）



絵入本『玉水物語』（京都大学貴重書デジタルアーカイブ）



名所『江戸名所図会』（日本古典籍データセット・国文学研究資料館撮影／味の素の文化センター蔵）

# マイクロコンテンツ化実践例： 「顔貌コレクション (顔コレ)」

- <http://codh.rois.ac.jp/face/>
- マイクロコンテンツの単位 = 美術作品に出現する顔貌
  - 日本の絵巻物を中心として顔貌を収集
  - メタデータを付与
  - 299作品・8,845顔貌 (2021年2月現在)

## 顔貌コレクション

### テーマ

顔貌 (8845)

### 性別

男 (6830) / 女 (1976) / 不明 (39)

### 向き

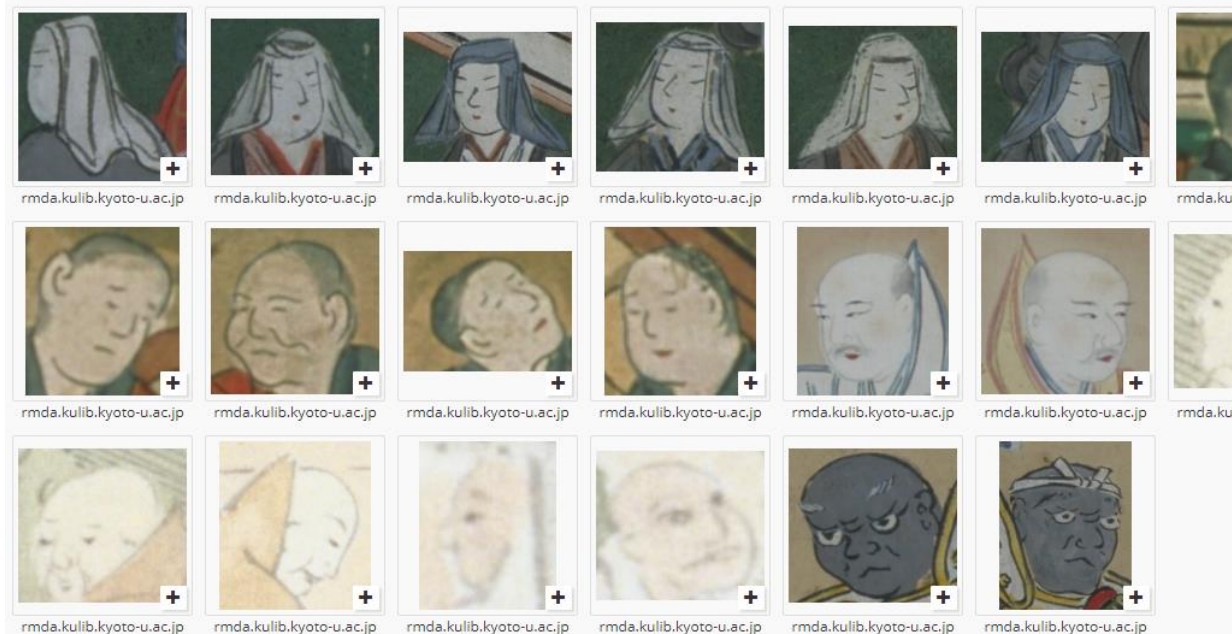
四分之三 (5176) / 横 (2771) / 正面 (102) / 不明 (11) / 後 (782)

### 身分

貴族 (3996) / 武士 (2877) / 庶民 (794) / 化身 (792) / **僧侶 (300)** / 動物 (20) / 神職 (10) / 不明 (56)

### 原典

検索結果 300 件 身分: 僧侶



< 1 2 ... 7 8 9 10 11 ... 14 15 >



# マイクロコンテンツ化実践例： 「顔貌コレクション（顔コレ）」

- 美術史研究（特に様式研究）への活用
- 描き方に現れる作者や工房の特徴（「様式」）を読み取りやすいため、顔貌表現は重要な素材



描き方をもとに、複数の作品から顔貌を集める  
『ふんせう』（慶應義塾大学）と『しつか』（国文学研究資料館）

# 「顔コレデータセット」

- <http://codh.rois.ac.jp/face/dataset/>
- Google Brain TokyoのYingtao Tianとの協力
- 「顔コレ」の顔貌を、機械学習に利用しやすい形式で提供するデータセット
- データセット化に伴いいくつかの工夫
  - 原典は宗教的内容も含むため、多様な価値観を尊重するガイドライン
  - 原典画像公開者への不利益を避ける工夫

# 顔貌コレクションを軸とした共同研究

- 日本中世絵巻における様式および男女の描き分けの研究
  - 高岸輝（東京大学美術史学研究室准教授）との共同研究
- マイクロコンテンツ化と分析に向けた研究対象
  - 『遊行上人縁起絵巻』の写本、『清浄光寺甲本』（以下「甲本」）
  - 時宗開祖一遍智真、二祖他阿真教の事績をまとめた多巻構成（全十巻）の絵巻
  - 非常に多くの登場人物が描かれている
  - 複数の絵師（絵師工房）の手によって描かれている

『遊行上人縁起絵巻 清浄光寺甲本』 (清浄光寺 遊行寺宝物館所蔵)  
画像提供 遊行寺宝物館



- 非常に多くの登場人物が描かれている
- 複数の絵師（絵師工房）の手によって描かれている

第三巻第一段「大衆に飲食を施す」

# 顔貌コレクションを軸とした共同研究

- 研究上の問い

1. 様式：複数の絵師の様式の峻別

- 複数絵師の様式（描き方の特徴）の違いを網羅的に確認できるか？

2. 図像：僧侶と尼僧の顔貌の分別

- 侍や貴族と異なり僧侶と尼僧のコード（衣装・髪型）は共通。ただし、男女の集団が明確に分かれているため、峻別可能
- 唯一の特徴である顔貌の描き分けは確認できるか？

# 僧侶と尼僧の共通性と、配置による区別



- 侍や貴族と異なり僧侶と尼僧のコード（衣装・髪型）は共通。ただし、男女の集団が明確に分かれているため、峻別可能
- では唯一の特徴である顔貌の描き分けは確認できるか？



「遊行上人縁起絵巻 清浄光寺甲本」 (清浄光寺・遊行寺宝物館所蔵)

# 遊行上人顔コレの構築

- 「顔コレ」の手法を使い、甲本に登場する顔をすべてマイクロコンテンツ化
- 顔貌を比較することで、様式・性差の描き分けを分析する

しかし…

- 全十巻の写本すべてから顔貌を抽出していくのは大きな作業負担になる

そこで…

# 機械学習による顔貌の自動認識

- 訓練データとして「顔コレデータセット」を利用
    - <http://codh.rois.ac.jp/face/dataset/>
  - 1. R-CNNを用いた顔貌検出器を、顔コレデータセットを訓練データとして学習させる
  - 2. 甲本画像をテストデータとして顔貌検出器に入力し、顔貌領域のXYWHを出力する
- 
- NII北本研究室インターンのAlexis Mermet (Ecole Polytechnique Fédérale de Lausanne 当時) の協力



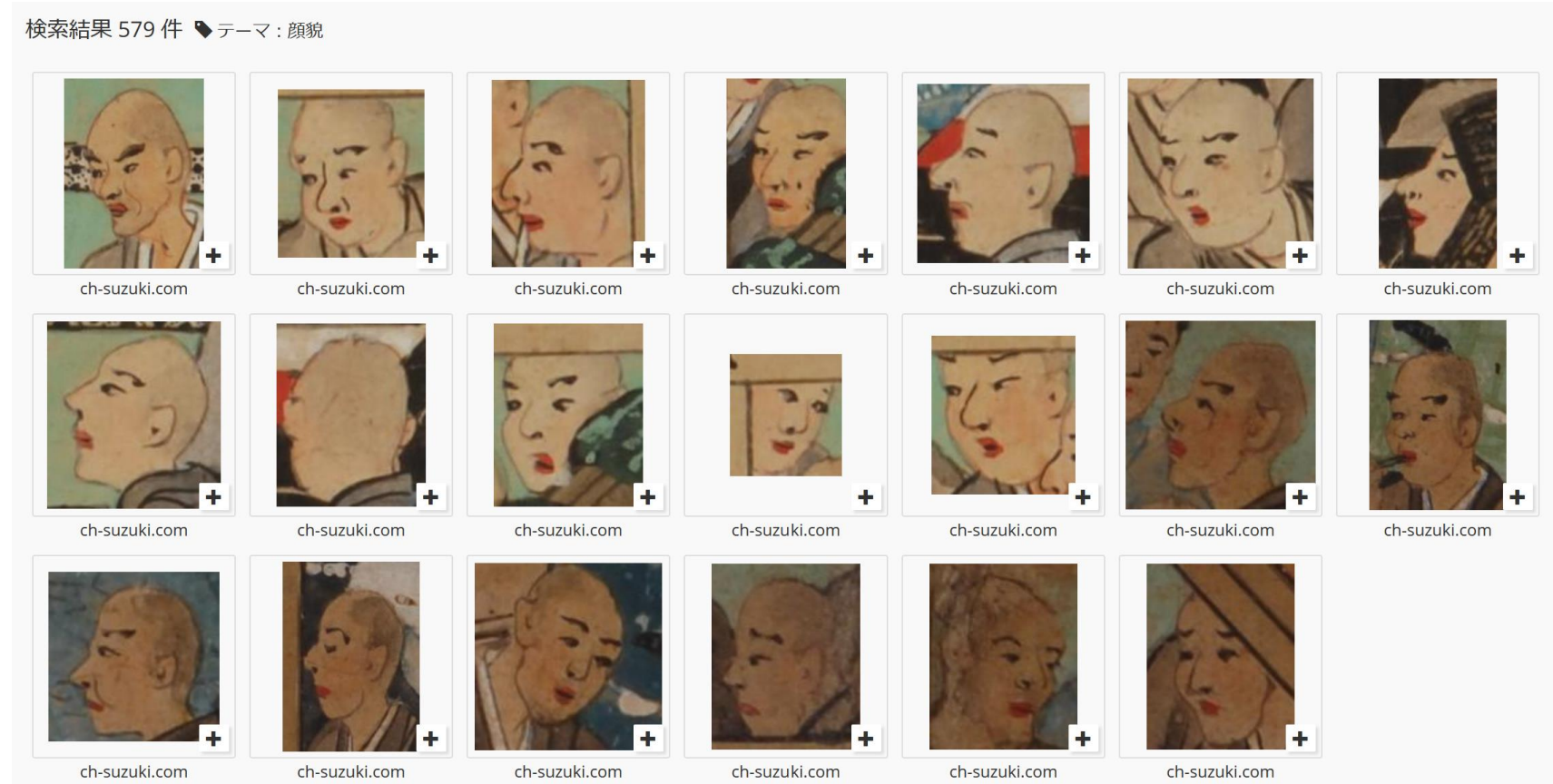
# 機械学習による顔貌の自動認識



群衆が重なり合って描かれた場面でも個々の顔貌を認識

# 分析の実践

- 機械学習の成果を使用して2,390顔貌を抽出
- メタデータを付与
  - 性別
  - 身分
  - 向き
  - 原典での登場場面
- 明確に判別できる僧侶と尼僧を抽出  
= 全28場面・579 顔貌



- これらをツールIIIF Curation Boardを利用し平面上に再整理
  - <http://codh.rois.ac.jp/software/iiif-curation-board/>



<http://codh.rois.ac.jp/software/iif-curation-board/demo/?curation=https://mp.ex.nii.ac.jp/api/curation/json/066ee1b7-f7cd-4260-b08a-71842b674d9a&lang=ja>

全顔貌を整理した状態

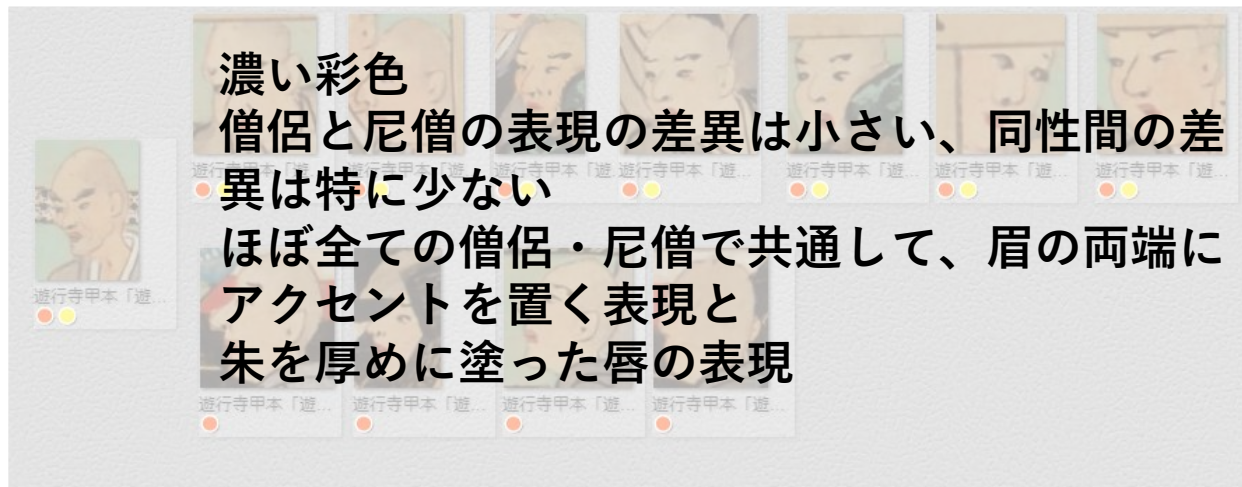
# 研究上の問いへの回答 1

様式：同一作品における複数の絵師の様式の峻別は可能か？

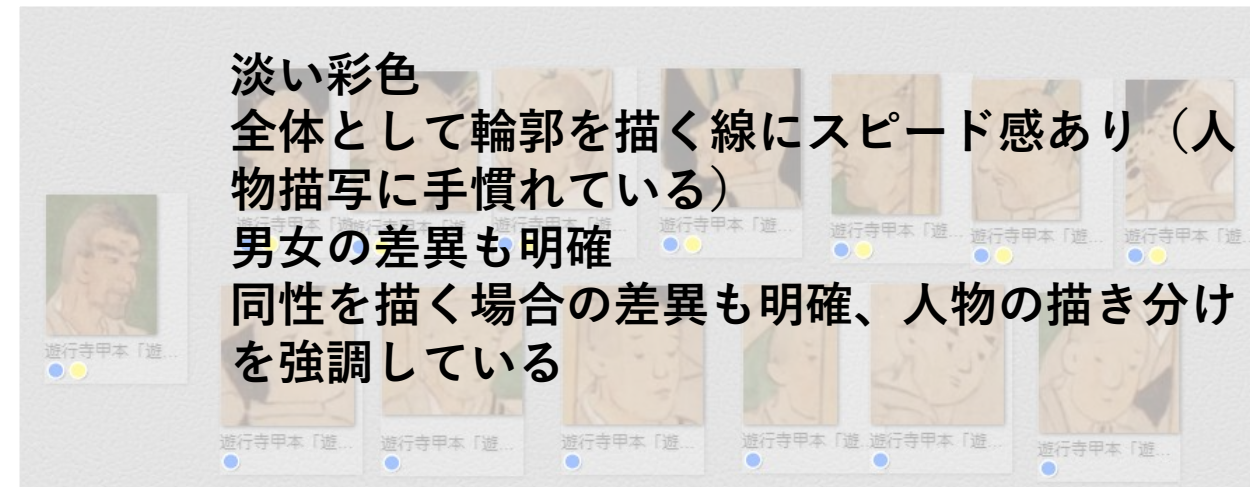
YES：網羅的に見ても、これまで言われてきた区分は説得性が高い

- 先行研究による分担制作 = 三人の絵師A・B・Cという仮説 (岩橋春樹『遊行上人縁起絵巻』 - 清浄光寺本について. 仏教芸術, 1989, vol.185, p.51-59.)
  1. 絵師A (579には該当の顔貌なし)

## 2. 絵師Bの顔貌



## 3. 絵師Cの顔貌

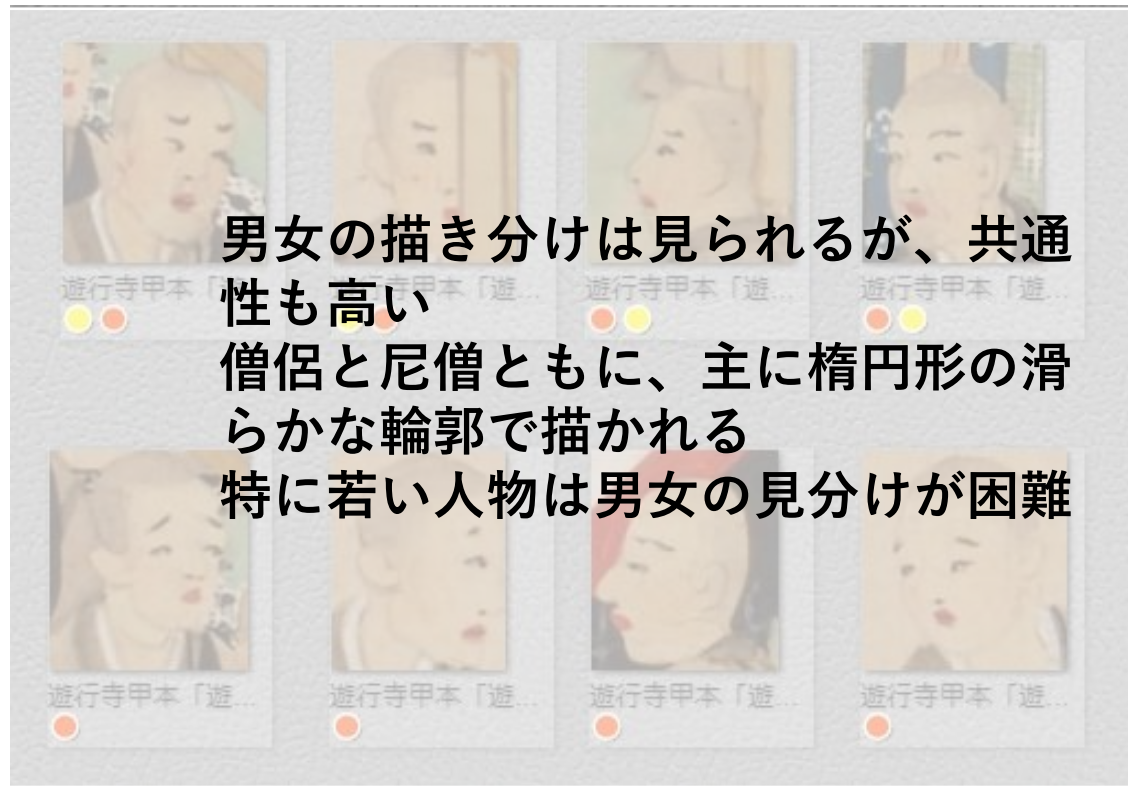


# 研究上の問いへの回答 2

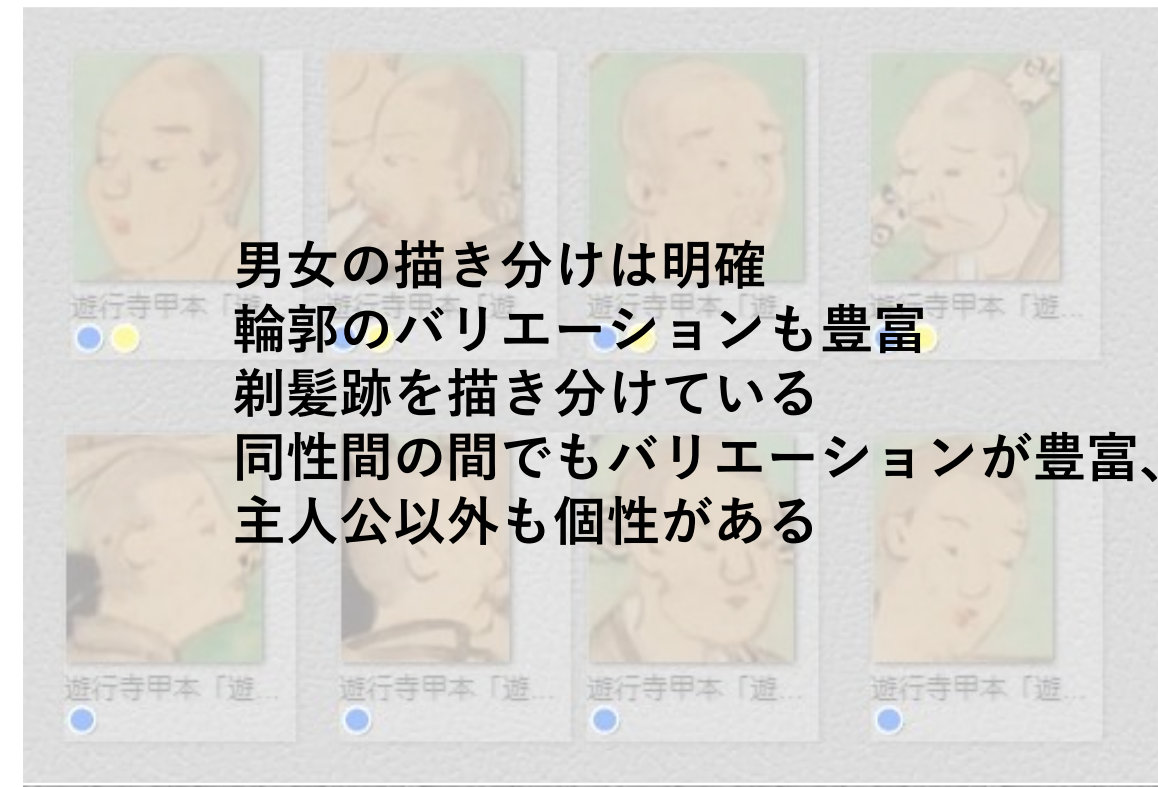
図像：僧侶と尼僧の描法の分別は可能か？

YES：顔貌による性差の描き分けが認められるが、絵師によって描き分けの差がある

1. 絵師Bの僧侶と尼僧



2. 絵師Cの僧侶と尼僧



# マイクロコンテンツを利用した分析 GM法

- マイクロコンテンツ化した人文学資料の分析手法として命名
  - 細部に注目した様式研究者Giovanni Morelliの名前とGazing Microcontentsの頭文字からの造語
  - 今後も引き続きブラッシュアップ
- ワークフロー

III Fに則ってデジタル化された資料

資料から部分を抽出・収集

メタデータの付与

横断検索・一覧化・活用

データセットの構築

分析

# 顔貌コレクションの別展開

- 「顔コレ」を利用した現代アートへの展開
- 国文学研究資料館「ないじえる芸術共創ラボ」  
<https://www.nijl.ac.jp/pages/nijl/>
- 参加アーティストである梁亜旋が、「顔コレ」をインスピレーション源として作品制作・ワークショップを実施
  - ワorkshop「古典籍の“面白い顔”が仮面になる！」
  - 梁亜旋展『古典から再構築へ』

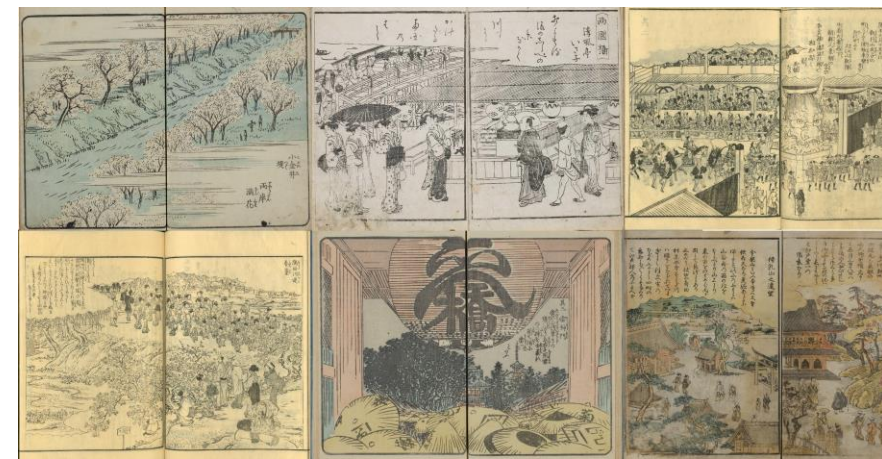


梁亜旋展『古典から再構築へ』



# 人文学資料マイクロコンテンツ化の展開

- データサイエンスに基づく人文学（人文情報学）研究の促進にむけて、様々な人文学研究資料へ応用可能
- 「江戸買物案内」「江戸観光案内」
  - <http://codh.rois.ac.jp/edo-shops/>
  - <http://codh.rois.ac.jp/edo-spots/>
- 江戸時代の資料から、商人や名所・観光地に関する画像情報を抽出
- 情報空間の中に江戸の要素を再構築していく
- 地理情報と結びつけることで、さらなる活用



# 今回の報告に関わる論文

- 鈴木 親彦・高岸 輝・本間 淳・Alexis Mermet・北本 朝展「日本中世絵巻における性差の描き分け－IIF Curation Platformを活用したGM法による『遊行上人縁起絵巻』の様式分析」『じんもんこん2020論文集』, pp.67-74, 2020
- 鈴木 親彦・北本 朝展・Yingtao Tian「顔コレデータセット：美術史研究分野における機械学習データセットの構築・公開の諸問題」『デジタルアーカイブ学会誌』4(2), pp.191-194, 2020
- Yingtao Tian, Chikahiko Suzuki, Tarin Clanuwat, Mikel Bober-Irizar, Alex Lamb, Asanobu Kitamoto, "KaoKore: A Pre-modern Japanese Art Facial Expression Dataset", arXiv:2002.08595.

# ご清聴ありがとうございました

第14回CODHセミナー

「IIF Curation Platform利活用レシピ100連発」

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

日時：2021年2月18日（木） 15:00-17:30

会場：オンライン（Zoom）

参加費：無料※ただし事前登録をお願いします

IIF Curation Platformの利活用事例における用途と、ICPの機能との関係を整理した「利活用レシピ」をご紹介します。

100連発で繰り出される多数のレシピから、ご自分のデジタルアーカイブ構築へのヒントをつかんで下さい。