

オープンデータと情報学で人文学はどう変わるか？

情報学と人文学のコラボレーションで進めるオープンサイエンス

どんな研究？

- 1. 情報学・統計学による新しい人文学
- 2. 情報学・統計学の人文学での活用
- 3. 情報・システム研究機構内にとどまらない、幅広い連携と発信
- 4. オープン化を軸とした研究全体の進展

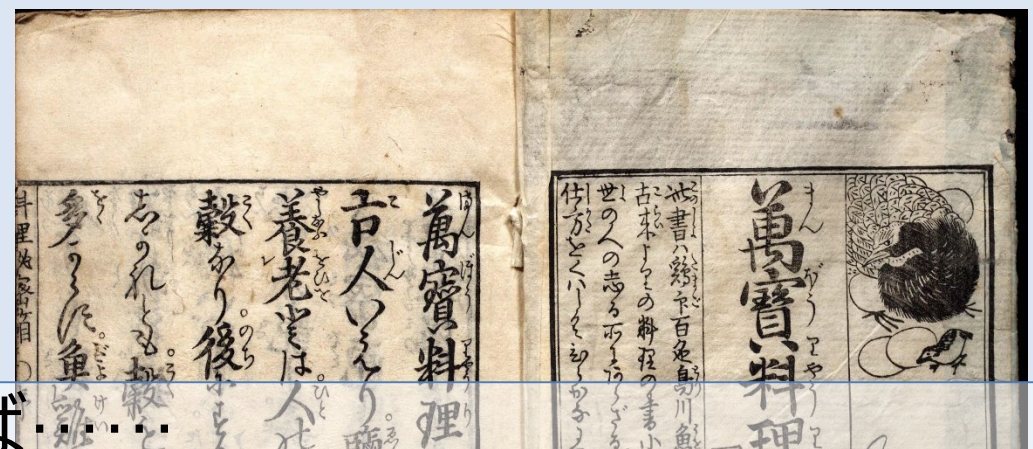
何がわかる？

- 1. データサイエンスに基づく「人文情報学」創生
- 2. 内容解析に基づく「ディープアクセス技術」
- 3. 組織の枠を超えた研究拠点の形成と、世界に向けての日本の人文知の発信
- 4. シチズンサイエンスやオープンイノベーションのモデル化

状況設定

四つの課題

- 1. データ利用基盤構築
- 2. 内容分析
- 3. 質的向上
- 4. オープン化

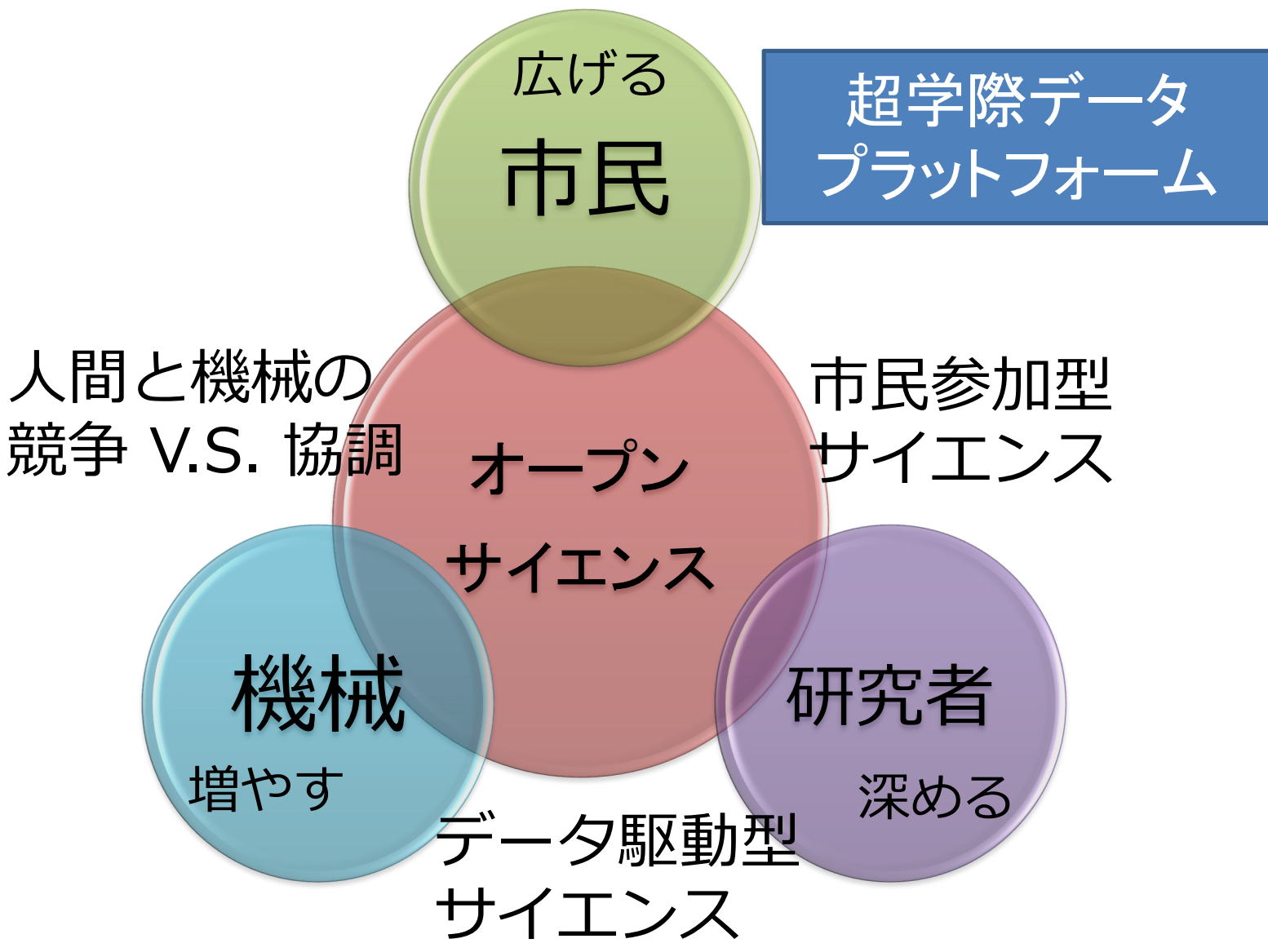


例えば……

- ・ AIでくずし字を解読
- ・ 統計数理による日本語コーパスの品質改善
- ・ デジタル人文資料のキュレーション
- ・ モバイルアプリを用いた市民科学

学問分野間の連携を超えた、超学際的データプラットフォームの必要性

市民・研究者・機械の全てが力を発揮できる環境の整備



研究内容

人文学オープンデータ公開・研究・利活用のプラットフォーム

基礎的なデータセットの構築

日本古典籍データセット

古典籍701点について、画像、書誌、タグなどをCCライセンスも明記してオープンデータ化。IIIFに対応した画像ビューアーもオープンソースとして構築。

国文学研究資料館・歴史的典籍NW事業との協力（国立情報学研究所）



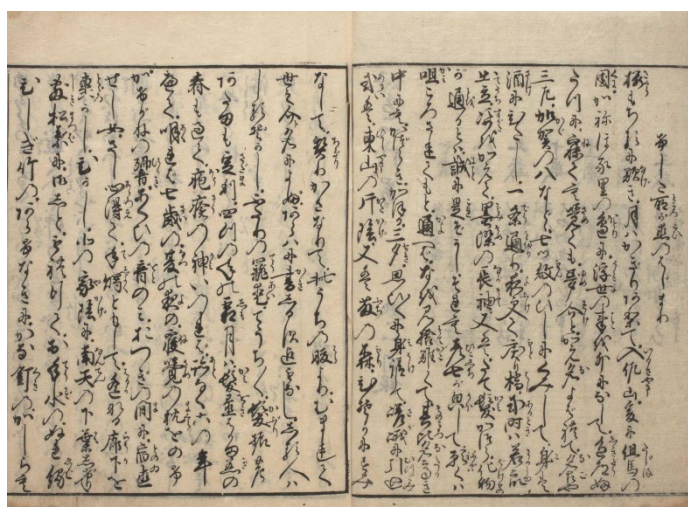
『源氏物語団扇画帖』 IIIF対応の高精細画像

字形を抽出

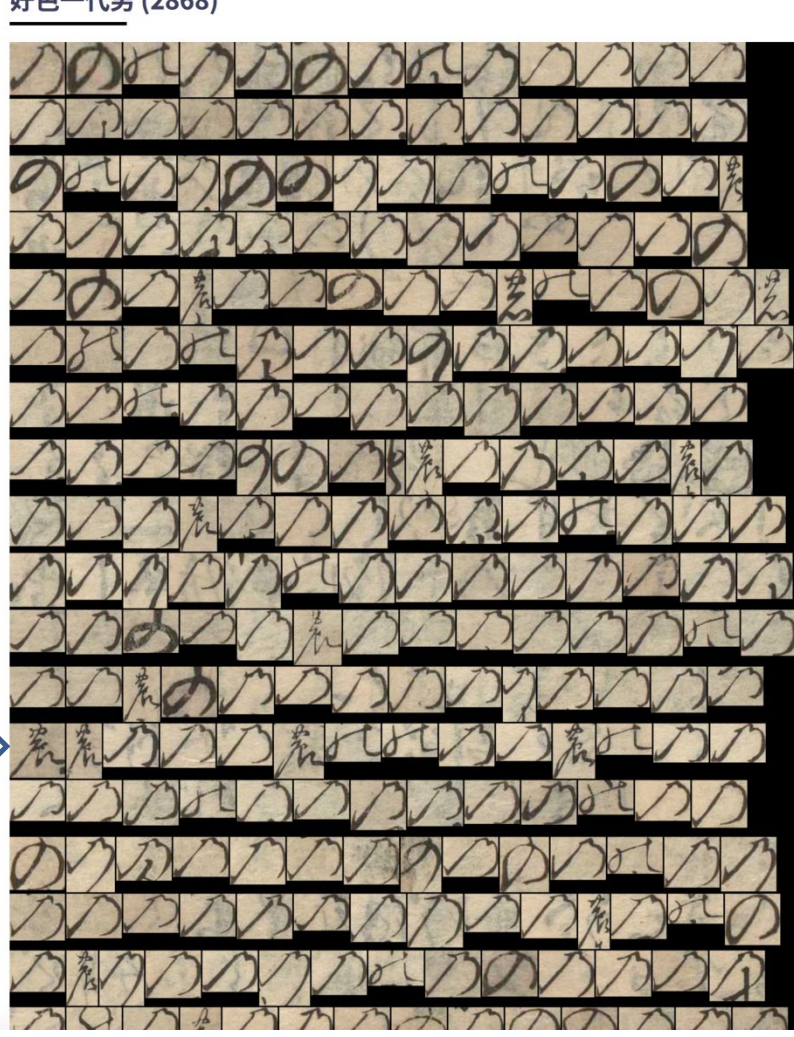
利活用を想定したデータセットの構築

日本古典籍字形データセット

日本古典籍データセットから画像データから切り取ったくずし字の字形データセット。2017年6月に現在は3,999種、403,242文字まで拡大。



『好色一代男』の全ての「の」



統計的大規模テキスト化

近代日本の文書画像資料を統計数理的なアプローチでテキスト化。膨大な作業が必要だった研究に、情報学のアプローチで再挑戦し、海外でも活用される日本研究・アジア研究の基盤を創出。

国立国語研究所および東京大学との協力（統計数理研究所）



既にデジタル化されている大量の近代人文学資料画像
雑誌『国民』『太陽』『東洋学芸雑誌』…

画像からテキストへの統計的復元



- ・ 検索可能
- ・ コーパス化
- ・ 量的な人文学研究

翻刻・現代語訳

江戸料理レシピデータセット

江戸の料理本を翻刻・現代語訳、実際に料理に使えるレシピとして公開。



江戸時代のレシピ本『万宝料理秘密箱 卯百珍』

レシピ概要	
項目	内容
食材分類	卵のみ
主な食材	卵
調味料	醤油
主な調理法	穴を空ける、漬ける（調味増）、煎貫（にぬき）
道具	針
主な食べ方	本汁、取巻、二の汁、刺身
翻刻テキスト	
手順	1 焼たまごの 新しさを 射にて焼の方へ すこしばかり穴をあけ 2 卵殻調味増へ 五日ほど漬おきて 3 取いたしよく水にて洗ひ 4 煎貫にすれば 中の黄味が 外へなり 白味が 中へ入ル 量を黄味返しといふ

レシピを翻刻・一部は現代語訳

データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター

Webサイトで様々なオープンデータを公開中

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



データセットを用いた参加型研究の展開

くずし字を機械で自動解読！

くずし字の画像と文字コードを組み合わせ、オープンデータセットとして提供、機械がくずし字を読むための学習データとして活用可能。

活用例：「PRMUアルコン」

コンピュータがくずし字を認識しUnicodeで出力するPRMU研究会のアルゴリズムコンテストが開催中。CODHが古典籍画像から抽出した字形データを提供。

応募開始：2017年5月31日

応募締切：2017年8月31日

詳細はアルコンHPにて

<https://sites.google.com/view/alcon2017prmu>

江戸料理をみんなで復活！創作！

「江戸料理レシピデータセット」プロジェクトとして、クックパッドと日本家政学会食文化研究部会が運営する「クックパッド江戸ご飯プロジェクト」に参加。

雑誌等メディアにも紹介され「つくレポ」もすでに複数の投稿が。

一是も紅煎貫たまごのごとくにかわを取て、ずいぶん熱湯へしばらく漬取出し、紙にまき、箸にても竹にてもかたを入レ（以下略）

① ゆで卵の殻を取り、熱い湯にしばらくつける。
② 湯から取り出し、紙に巻き、箸や竹の型に入れ（以下略）

ゆで卵を作り、卵の殻をむきます。卵をむいた卵を熱湯にししばらくつけます。卵を湯から取り出し、クッキングシートを巻きます（なくても出ます）。お箸を使って形をつけます。箸の上からゴムでくります。よく冷ました後、食紅で色をつけます。箸に切ります。



江戸料理レシピデータ構築協力：
合同会社AMANE、グルメ クッキング ファクトリー、ピッツェリア ジオ・ピッポ

クックパッド「江戸ご飯」
<https://cookpad.com/kitchen/14604664>

研究開発人材募集中！

人文学オープンデータ共同利用センターでは、画像・言語解析への挑戦に興味ある**研究員を募集**しています。

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

これからの取り組み

ディープアクセス技術

近世(古典籍)+近代(明治～戦前頃の活字本)の内容にアクセスするための研究開発。

研究コミュニティ

シルクロード研究・大量画像による美術史研究など、人文学研究コミュニティとの連携を強化。

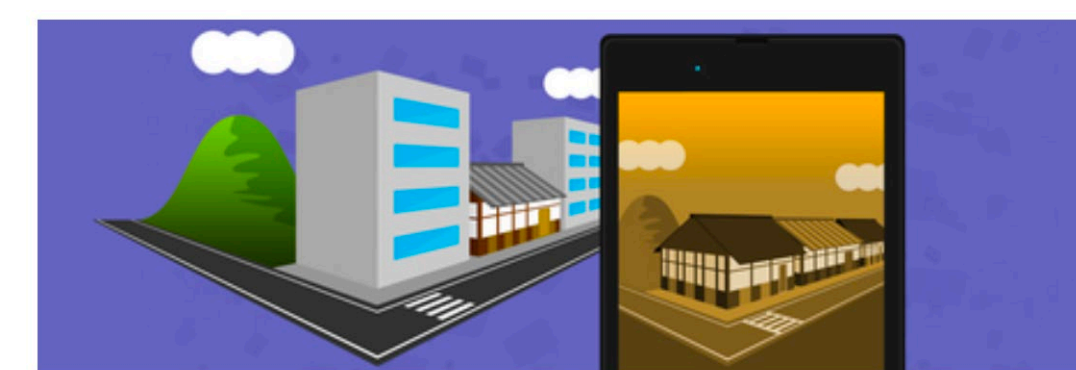
人文学研究データ基盤

人文学研究データキュレーション手法の創生によるデータ活用の促進。データ駆動型人文学の基盤となるプラットフォーム形成。

その他にも



デジタル・シルクロード



メモリープラットフォーム
メモリーグラフ



GeoNLP



IIIF Curation Viewer



華北交通アーカイブ



人文学研究データリポジトリ

ぜひWebサイトをご覧ください