

# くずし字学習、文字認識に朗報！ 日本古典籍字形データセット 8万字を公開

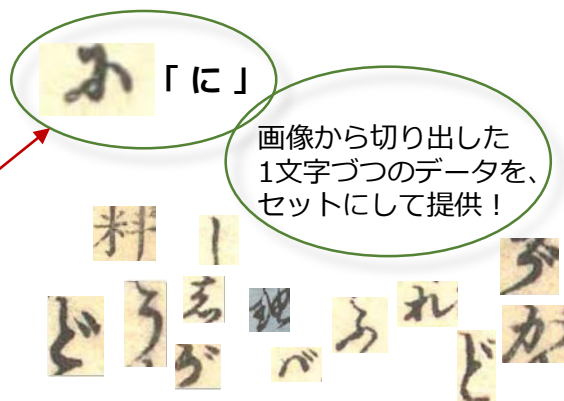
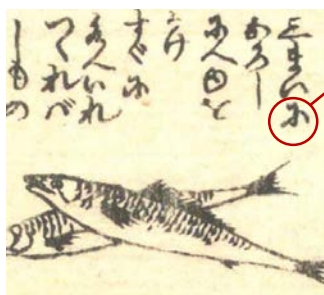
## ● 古典籍画像データのさらなる活用を目指して - 字形約80,000字をCC BY-SAで公開

国文学研究資料館は、デジタル撮影された当館所蔵古典籍を翻刻する過程で生まれたくずし字文字の情報等を、オープンデータセットとして、人文学オープンデータ共同利用センター（準備室）より、研究用に提供いたします。

1文字ずつの画像、文字座標データなどを1セットにして提供します。古典籍に書かれた文字の機械学習用データなどへの活用が出来ますので、そこからくずし字解読の効率化などが期待されます。今後、公開する字形データを拡大していく計画（今年度中に合計約40万字を予定）です。ご期待ください。


## ● 字形のもとになった古典籍は料理本8作品分

- ◆ 当世料理
- ◆ 万宝料理秘密箱
- ◆ 膳部料理抄
- ◆ 料理物語
- ◆ 日用惣菜俎不時珍客即席庖丁
- ◆ 料理方心得之事
- ◆ 新編異国料理
- ◆ 料理秘伝抄



これが8万字分！

## ● 日本古典籍字形データセットについて

公開日	2016年11月17日（木）
公開サイト	<ul style="list-style-type: none"> <li>・ 国文学研究資料館 <a href="http://www.nijl.ac.jp/pages/cijproject/data_set_list.html">http://www.nijl.ac.jp/pages/cijproject/data_set_list.html</a></li> <li>・ 情報・システム研究機構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（準備室） <a href="http://codh.rois.ac.jp/char-shape/">http://codh.rois.ac.jp/char-shape/</a></li> </ul>
データ数	上記8点の資料に書かれた、86,176文字
データ概要 ※①～④のデータを1セットにして8万字を公開	<ol style="list-style-type: none"> <li>① 原本補正画像データ：日本古典籍データセットで公開する画像に対して、翻刻作業を容易にするための前処理として、見開き画像を分離するとともに、回転させて正立させるという処理を加えた画像です。</li> <li>② 文字座標データ：原本補正画像データ上で文字を取り囲む長方形の座標（XYWH）、文字のUnicodeコードポイント、ブロックID、文字IDを記録したものです。</li> <li>③ 字形画像データ：「原本補正画像データ」に「文字座標データ」を適用して切り抜いた画像であり、文字種ごとに字形を閲覧しやすくするために提供するものです。</li> <li>④ 作業報告文書：作業で読めなかった文字に関する情報や、その他の注意事項を記したドキュメントです。</li> </ol>
利用条件	クリエイティブ・コモンズ・ライセンス表示－継承4.0 <b>CC BY-SA</b> この条件に同意される方であればどなたでもご利用頂けます。 <a href="https://creativecommons.org/licenses/by-sa/4.0/deed.ja">https://creativecommons.org/licenses/by-sa/4.0/deed.ja</a> 
お知らせ	古典籍画像の利活用を考えるため、公開にあわせてアイデアソンを開催。 ◆ 歴史的典籍オープンデータワークショップ（アイデアソン） ◆ 2016年12月9日（金）午後、国文研（東京都立川市）にて ◆ 申込サイト <a href="http://ptix.co/2fJCE2S">http://ptix.co/2fJCE2S</a>

## ■ プレスリリース ■

メディアの皆様からのお問い合わせ先：国文学研究資料館

【本件の問い合わせ先】

国文学研究資料館 古典籍共同研究事業センター事務室

古典籍共同研究係

e-mail : cijinfo@nijl.ac.jp

TEL : 050-5533-2988 FAX : 042-526-8883

大学共同利用機関法人 人間文化研究機構 国文学研究資料館について

(東京都立川市, 館長：今西祐一郎)

国文学研究資料館は、国内各地の日本文学とその関連資料を大規模に集積し、日本文学をはじめとする様々な分野の研究者の利用に供するとともに、それらに基づく先進的な共同研究を推進する日本文学の基盤的な総合研究機関です。

2014年度からは、人文系では初の大規模学術フロンティア促進事業「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」がスタートしており、国文研が中心となって国内外の大学等と連携し、日本語の歴史的典籍のデジタル化と、国際共同研究ネットワークの構築を推進しています。

【国文学研究資料館】

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

【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ついて】

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

メディアの皆様からのお問い合わせ先：国立情報学研究所,  
人文学オープンデータ共同利用センター

【本件の問い合わせ先】

大学共同利用機関法人 情報・システム研究機構

国立情報学研究所 総務部企画課 広報チーム

e-mail : media@nii.ac.jp

TEL : 03-4212-2164

情報・システム研究機構・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について

(東京都千代田区, 準備室長：北本朝展 (国立情報学研究所))

2016年4月1日に、情報・システム研究機構データサイエンス共同利用基盤施設にて活動をスタート。NIIコンテンツ科学研究系准教授の北本朝展が室長となり、NIIと統計数理研究所との共同研究、および国内外の人文学研究機関との連携を軸に、人文学オープンデータの共同利用という課題に取り組む。来年4月にセンター化予定。

【人文学オープンデータ共同利用センター準備室】

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