

「理系と文系」の壁を越えて広がるオープンサイエンス

北本 朝展 (KITAMOTO Asanobu)

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

国立情報学研究所 コンテンツ科学研究系

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

自己紹介



- 研究分野は**情報学**。画像処理や画像データベースの研究が出発点。
- データ駆動型サイエンスに発展し、**気象情報**、**地球環境情報**、**人文科学情報**などの分野で研究。
- 最近はおープンサイエンスにも関わる。

オープンサイエンスとは？

- 「オープン」という言葉を梃子にして、より良いサイエンス（研究）に変える。
- 「よりオープンに」という方向性を共有する活動を、一語で束ねることの価値。
- 個々の活動ごとに「オープンサイエンス」の実態は異なり、単一の定義は困難。
- 大同団結？ 同床異夢？ 個々の活動を超える新しい目標を示せるかが問われる。

はじめに

デジタル台風

http://agora.ex.nii.ac.jp/digital-typhoon/

デジタル台風：台風画像と台風情報

ホーム Earth デジタル台風 English

お知らせ：ご意見：登録：動画：キッズ：RSS / Atom / Media RSS：OpenSearch
ケータイ版：PDA版：Google Earth版：iPod版：Twitter版：Facebook版：デジタルフォトフレーム版：アプリケーション
ソーシャル台風：台風前線：台風画像：台風空想：アイフーン：ツイフーン：台風なう！：台風メモリーズ：ふってきったー

リアルタイム台風情報 更新日時：2013年10月07日16時49分(JST)

台風201323号	台風201324号
MTS213100706	MTS213100706
201323 (WNP)	201324 (WNP)
(N27.0, E118.0)	(N26.7, E128.7)
1000 hPa / 0 kt	935 hPa / 100 kt

台風発生数
現在 = 24個 (台風経路図) (台風進路予想図)
平年 = 19.8個 (1951-2010)

最新台風情報

- 台風ニュース・ウェブログ
 - 2013年台風23号 (フィートウ | FITOW) (2013年10月05日)
 - 2013年台風24号 (ダナス | DANAS) (2013年10月07日)
- 台風ニュース・トピックス
- 地域情報ポータル
- アメラス (アメダス・ランキング)
 - 降水量マップ：風向・風速マップ：最新画像：イベント抽出
 - レーダー：Google Maps版：最新画像
- 災害情報データベース
- その他のニュース

台風データベース

静止気象衛星画像 (地球)

1. メタデータによる検索

- 日時・シーズンで検索
- 名前・番号で検索
- 地図で検索
- 地名 (緯度・経度) で検索
- 最低中心気圧で検索
- 最大風速で検索
- 日付で検索
- 活動カレンダー (存在確認)
- 台風タイムライン

北西太平洋：南西太平洋

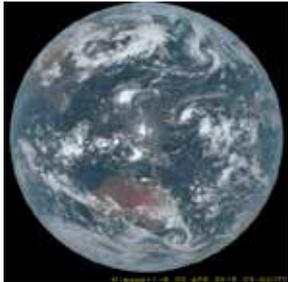
2013年10月7日15時 (JST)

www.digital-typhoon.org
2013-10-07 06:00 UTC

- 台風に関する「あらゆる情報」を集約。
- 最新状況の把握と過去アーカイブの検索がシームレスに接続。
- 利用実績は年間2千万ページビュー以上。
- 研究者のみならず、一般市民も多数利用。

DIAS : 地球環境データ

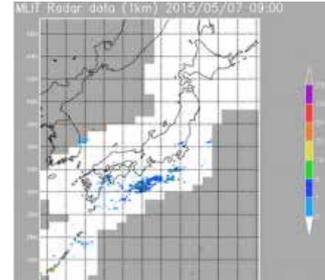
Data Dissemination



Himawari-8 Satellite



Weather Forecast GPV

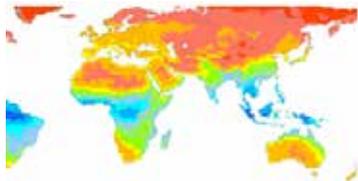


Radar Data

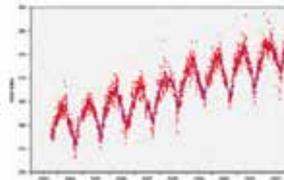


River Telemeters

1. Climate



CMIP5 Model



Visualization Tools

2. Water



Dam Control



Water Management

3. Agriculture

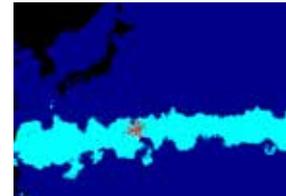


Potential of Rice Crops after Climate Change

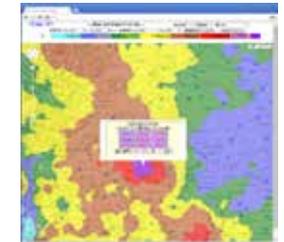
4. Biodiversity



Citizen science-based observations



Fish eggs and growth distribution



Accumulated radar rainfall

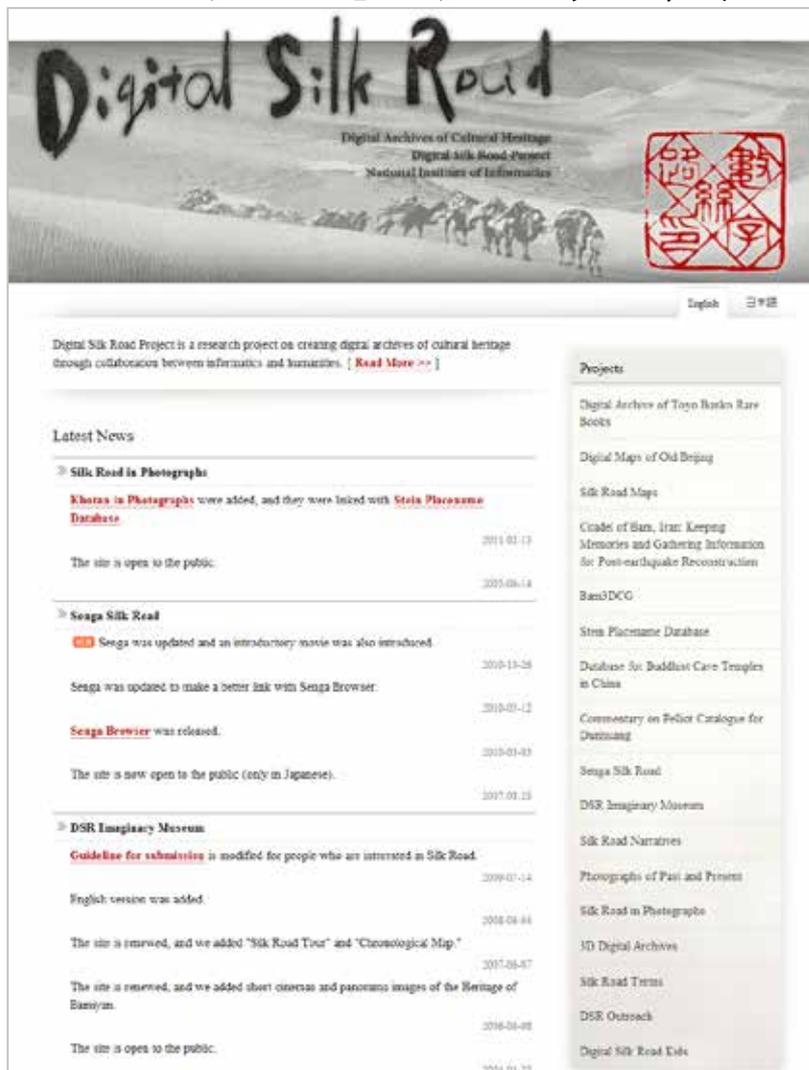


エレクトリカル・ ジャパン

1. 日本最大の位置情報つき発電所データベース・検索機能（2018年2月現在約9,200件）
2. リアルタイム電力需給データの網羅的アーカイブ
3. 電力統計データの多角的なビジュアライゼーション
4. 発電所の歴史のアニメーション
5. 電力データと気象データ、為替データなど異種データの統合可視化
6. 米国のデータも同様の方法で可視化

デジタル・シルクロード

<http://dsr.nii.ac.jp/>



- 人文情報学プロジェクト：情報学と人文学の共同研究によるデジタル人文学。
- データベースや研究リソース等は、ウェブで一般に公開。
- データの内容を分析する技術も研究開発。

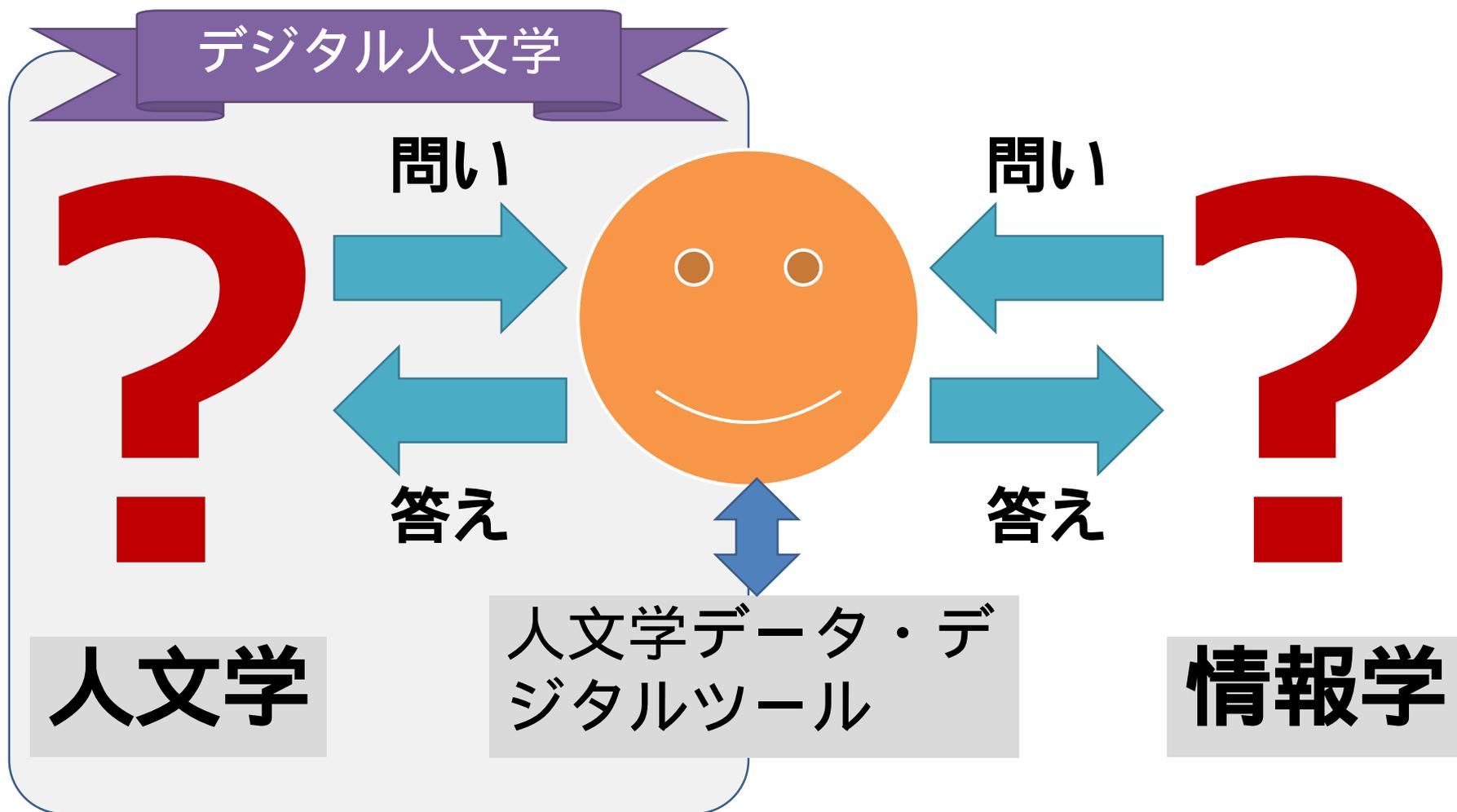


人文学オープンデータ 共同利用センター

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

- 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内に、2017年4月1日に正式に発足。センター長：北本朝展。
 1. **情報学・統計学の技術を用いて、人文学の研究を革新する。**
 2. 人文学のデータを用いて、情報学・統計学の研究を革新する。

学際領域と同床異夢



データ駆動型
サイエンス

深める

研究者

増やす

機械

オープンサイエンス

市民参加型サイエンス
(シチズンサイエンス)

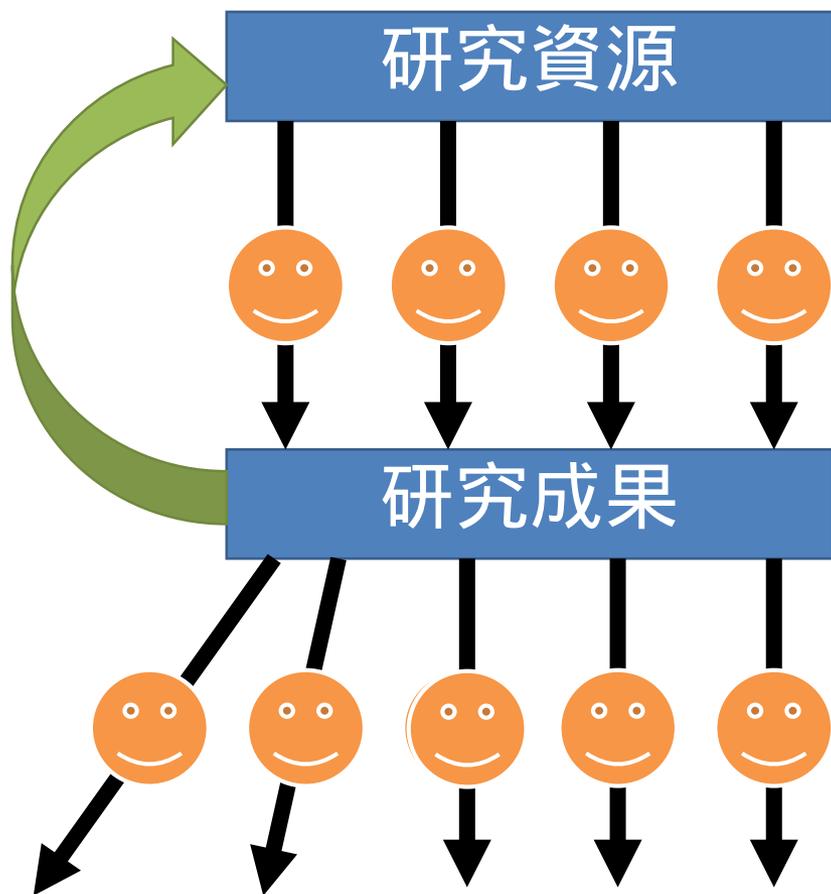
人間と機械の競争
vs. 人間と機械の協調
(チームング)

市民

オープンサイエンスの共創
トライアド

広げる

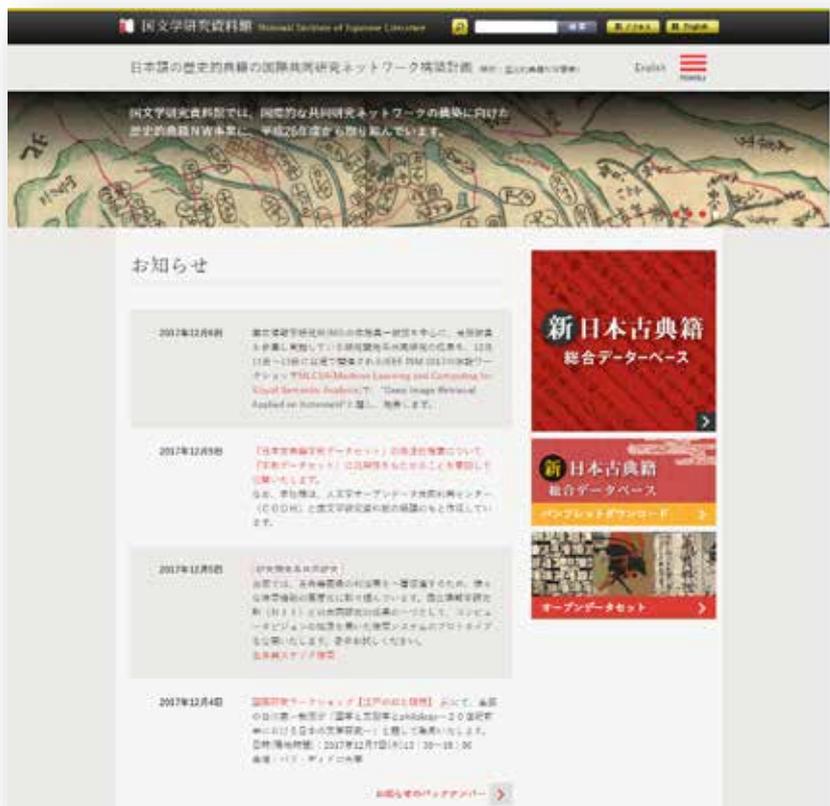
研究サイクルのオープン化



- **研究の資源**となるデータやソフトウェアを共有。
- オープンな資源をそれぞれの目的に活用。
- **研究の成果**となる成果物を共有。
- オープンに学びあいつつ**仲間を広げる**。

機械のためのオープン データ

NIJL-NWプロジェクト



- 「日本語の歴史的典籍」約30万点を画像データ化。
- 日本文化に関する、初の(?)ビッグデータ公開事例。
- CODH / NIIは、オープンデータ公開やデータの利活用のテーマで、国文研と協働。

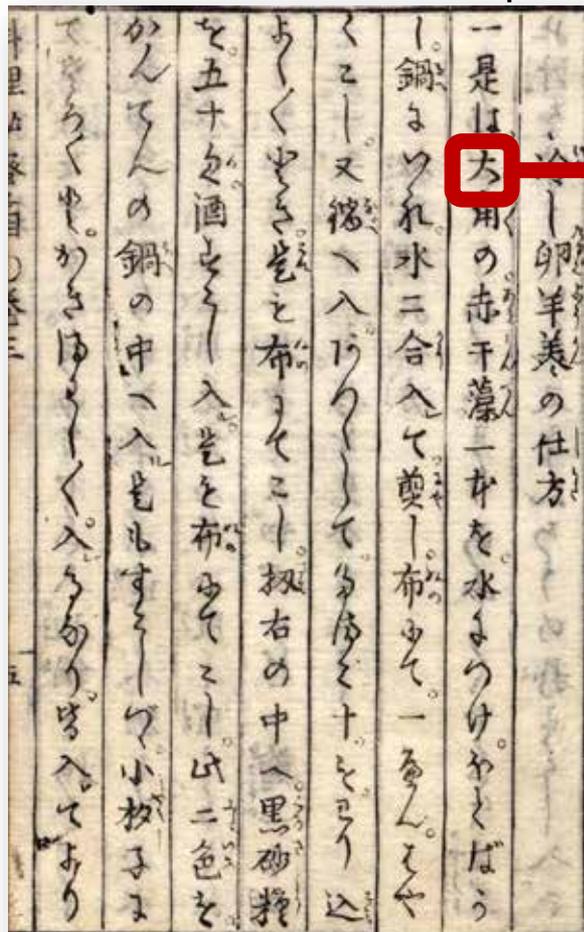
<http://www.nijl.ac.jp/pages/cijproject/>

ディープアクセスと データの構造化

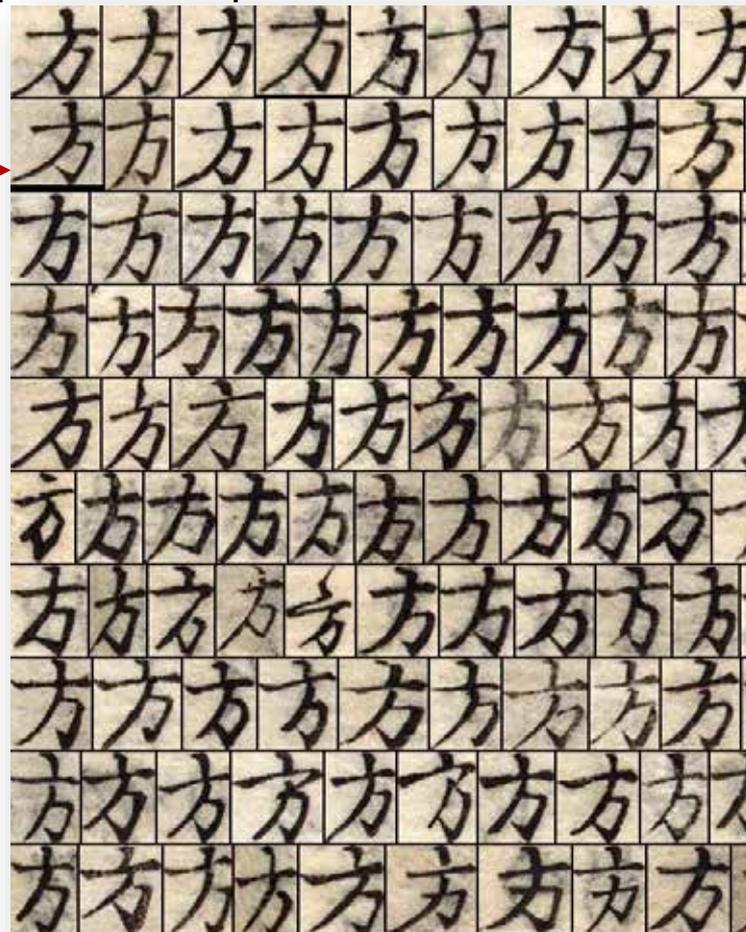
- **ディープアクセス**：古典籍のメタデータに加え、内容分析を用いて直接アクセス。
 1. **日本古典籍字形データセット**：古典籍の文字情報と座標を構造化。
 2. **江戸料理レシピデータセット**：料理本の記述を調理可能なレシピに構造化。
 3. **武鑑全集**：人物情報や組織情報、地理情報などを構造化。

機械のためのオープンデータ

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



日本古典籍データセット
(国文研所蔵)



日本古典籍字形データセット
(国文研所蔵・CODH加工)

日本古典籍字形データセット

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

人文学オープンデータ共同利用センター学術室
Center for Open Data in the Humanities

日本古典籍字形データセット

日本古典籍データセットで公開するデジタル化された古書籍を対象に、翻刻テキストを作成する過程で生まれるくずし字の座標情報などを、機械のための学習データや人間のための学習データとして提供します。

字形データセットの一覧を見る (機械のための学習データ)

文字種ごとの字形一覧を見る (人間のための学習データ)

2016年11月現在、日本古典籍字形データセットの規模は、日本古典籍データセットの8点の画像データから切り取ったくずし字1,521文字種の字形データ86,176文字です (文字種のリスト)。2016年度中には約40万字に規模を拡大する計画ですので、どうぞご期待下さい。

データ概要

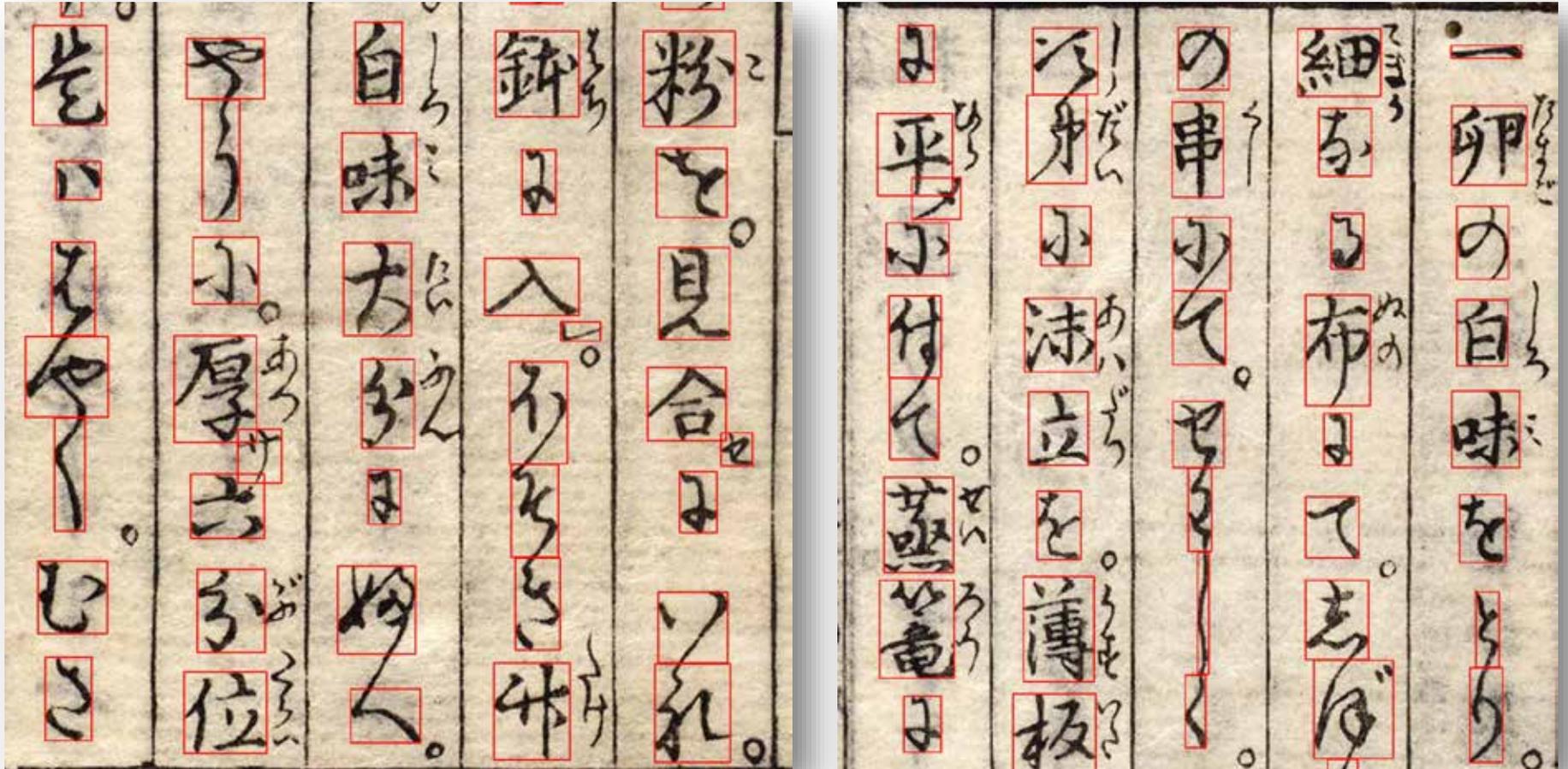
原本補正画像データ	日本古典籍データセットで公開する画像に対して、翻刻作業を容易にするための前処理として、見開き画像を分割するとともに、回転させて正立させるという処理を加えた画像です。
文字座標データ	原本補正画像データ上で、文字を取り囲む長方形の座標 (X,Y,W,H)、文字のUnicodeコードポイント、ブロックID、文字IDを記録したものです。
字形画像データ	「原本補正画像データ」に「文字座標データ」を適用して切り抜いた画像であり、文字種ごとに字形を閲覧しやすするために提供するものです。
作業報告文言	作業で成りなかつた文字に関する情報や、その他の注意事項を記したドキュメントです。

ライセンス

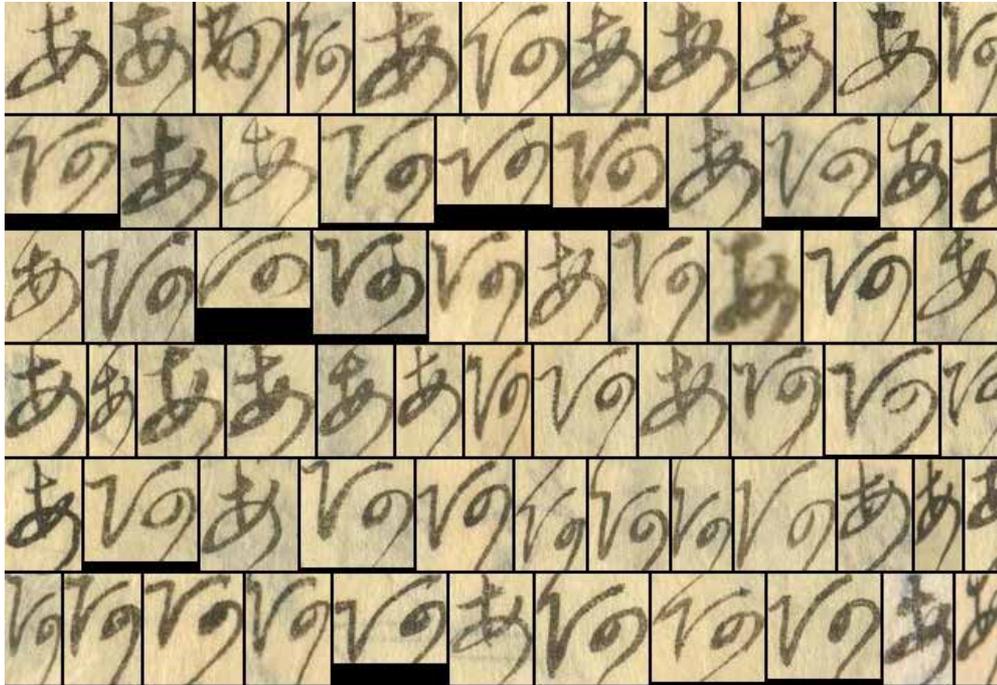
「日本古典籍字形データセット」(国文学研究資料館所蔵・情報・システム研究機構データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター加工)はクリエイティブ・コモンズ 表示・継承 4.0 国際 ライセンス (CC BY-SA) の下に提供されています。

- 2016年11月17日に公開。
- CC BY-SAライセンスのオープンデータ。
- 翻刻の副産物である文字の座標情報を、人間と機械を賢くするためのデータとして利用。

字形画像データ



人間のための学習データ



変体仮名「あ」

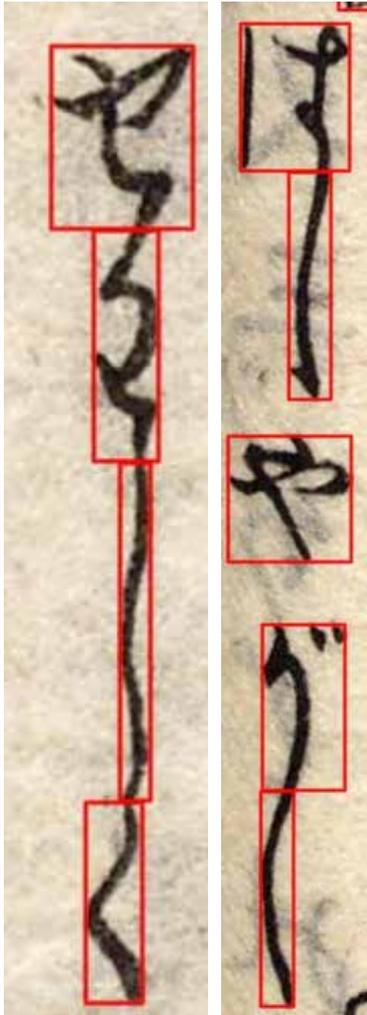
- 字形を目で見て確認できる。
- **学習アプリの素材にもなる。**
- くずし字が読める人が増える→データの利活用が進む→くずし字がより広まる。

機械のための学習データ

文字種	文字数
に	15,982
の	14,337
し	13,386
て	12,976
と	11,068
を	9,208
は	8,725
り	8,656
か	8,439
な	7,892
3,999文字種	403,242文字

- 現在403,242文字、近日中に70万文字以上に。ただし出現頻度が少ない文字もあり、十分ではない。
- **座標情報**を使えば、個別の文字より大きい単位で認識することも可能。
- **変体仮名の字母**は区別していない点に注意。

分割と認識の深い関係



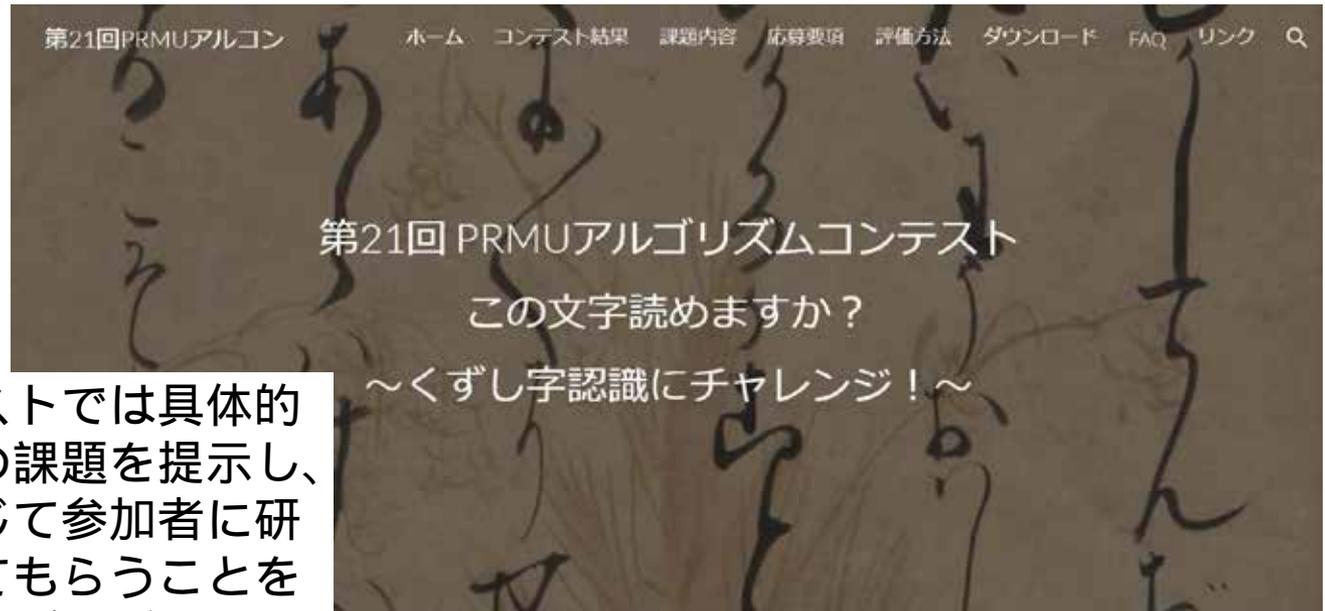
1. 文字を認識するには、文字を分割しなければならない。
2. 文字を分割するには、文字を認識できていなければならない。
3. 分割の自動化が必要かどうかで、問題の難易度は大きく異なる。
4. 分割という難問を回避する文字認識手法も研究が進んでいる（はこだて未来大学の寺沢氏による）

くずし字チャレンジ！

1. **機械学習（ディープラーニング）**：近年の技術の進展に伴い、難問に再びチャレンジする機運が生まれてきた。
2. **コンペティション型研究**：共通のデータセットに対して、様々なアルゴリズムを提案し、性能を競い合う。
3. **くずし字チャレンジ！**：文字認識やレイアウト解析などの問題を設定し、機械がくずし字を読む技術を競う。

PRMUアルゴリズムコンテスト

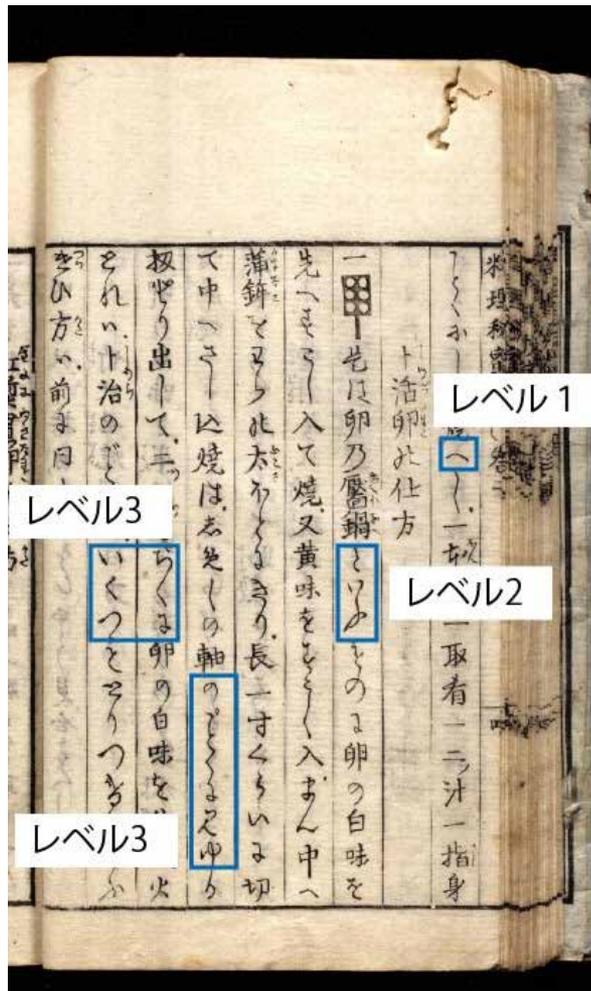
電子情報通信学会
パターン認識・メ
ディア理解研究会
(PRMU)



アルゴリズムコンテストでは具体的なパターン認識分野の課題を提示し、課題解決の過程を通じて参加者に研究の面白さを体験してもらうことを目指しています。アルゴリズムコンテストで提示する課題は代表的・基礎的な研究課題を取り上げますので、研究室に配属された学生が**パターン認識・メディア理解の技術を学ぶ教材として、また、授業の課題としても最適です。**

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

コンテストの課題



1. 課題の難易度

- レベル1は1文字
- レベル2は縦方向の3文字
- レベル3は縦横方向の3文字以上の文字

2. 認識対象の文字

変体かな50種程度、漢字含まず。

3. 評価方法

- くずし字認識の精度
- アルゴリズムの独創性（既存手法の単純適用は低評価）
- データに対するロバスト性

コンテストの結果

The IAPR Best Paper Award of HIP 2017 is presented to

Hung Tuan Nguyen, Nam Tuan Ly, Kha Cong Nguyen, Cuong Tuan Nguyen and Masaki Nakagawa

for their paper entitled

Attempts to recognize anomalously deformed Kana in Japanese historical documents

We congratulate the authors for their outstanding work, which is breaking ground for the challenging problem of Kana recognition in historical Japanese documents, thoroughly investigates state-of-the-art methods at different recognition levels, and promotes the use of open data for historical document imaging and processing.

Andreas Fischer, Angelika Garz, Kengo Terasawa, and Bill Barrett
IAPR Best Paper Award Committee



<http://events.unifr.ch/hip2017/>

- コンテストで1位の東京農工大学 中川研究室 チームは、ベトナム人が主力。
- コンテストの成果が、国際会議でもベストペーパー賞を獲得。
- 今年もコンテストを開催したい。もっと国際的な規模で。

歴史的文書の文字認識

1. **人文学からのニーズはむしろ高まりつつある。** マイナー言語のOCRも課題。
2. 完璧に読めればいいが、まず欲しいのは検索で、**最後は人間が読めばよい。**
3. **大規模学習データセットの構築が課題。** オープンデータはまだ少ない。
4. ユーザが限られるため、商用化は難しく、むしろ**学术界の貢献に期待が高い。**

フォーマットの共通化

「字形データセット」データ作成（基本仕様）

「日本古典籍字形データセット」の各データについて、国文学研究資料館（以下「当館」という。）がどのような仕様で発注しているか、仕様書の内容を提供することにより、「字形データセット」に汎用性をもたせていきたいと考えています。

業者に発注する場合はもちろんのこと、個人でデータを作成する場合の作成ルールとして、ご活用ください。

本基本仕様のライセンスは、クリエイティブ・コモンズ表示-継承 4.0 国際ライセンス (CC BY-SA) です。「国文研字形データセット仕様」を用いた旨お書きください。

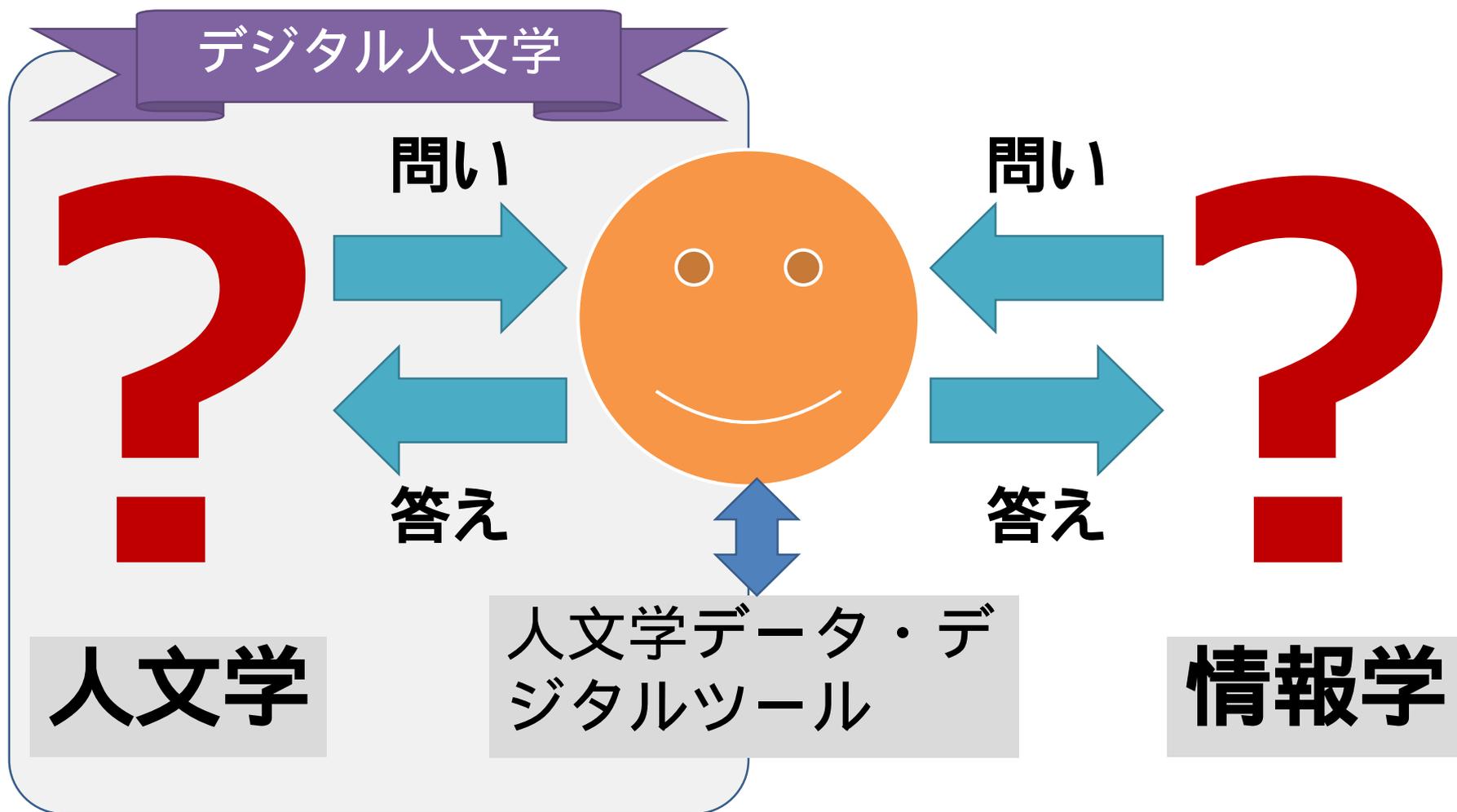


朱書き部分は当館の部署、規程等に関わる部分及び本基本仕様に対する注（\$表示）ですので、各発注者の事情に合わせて変更ないし削除願います。

http://www.nijl.ac.jp/pages/cijproject/data_set_list.html#section_004

- 国文学研究資料館が、共通的に使える基本仕様をオープン化。
- この仕様に沿ってデータを作れば、みんなが学習データを簡単に共有できる。

学際領域と同床異夢



市民のためのオープン データ

オープンサイエンスの要素

透明性

オープンアクセス

共有

オープンピアレビュー

オープンデータ

研究の再現性・
透明性・研究
データ保存

研究データ
データ出版
データリポジトリ

市民科学・クラウドファンディング

コラボレーション・オープンイノベーション

超学際研究

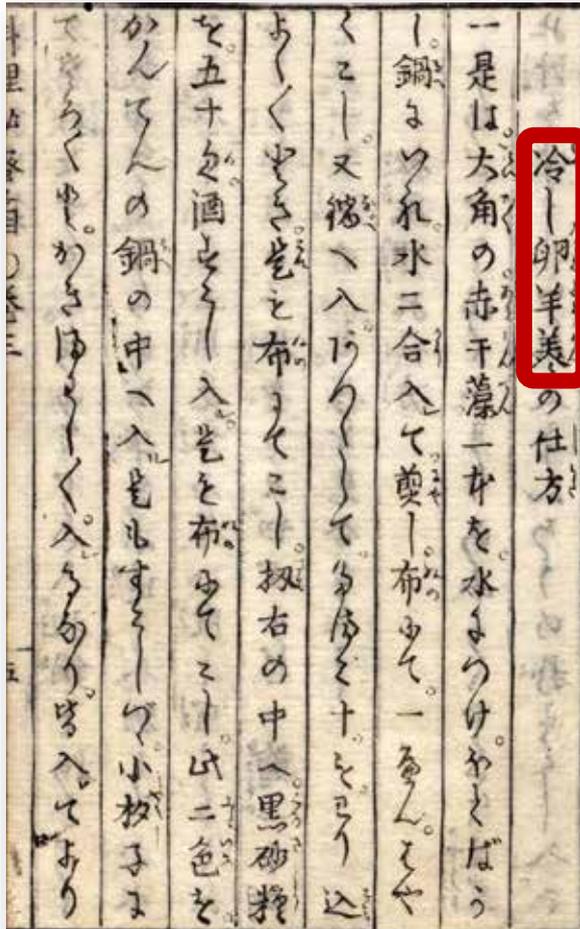
参加

協働

メタ研究 = 研究（システム）に関する研究

市民のためのオープンデータ

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



日本古典籍データセット
(国文研所蔵)



江戸料理レシピデータセット
(CODH制作)
日本古典籍データセット
(国文研所蔵)を翻案

じんもんそん2015

国文学研究資料館 National Institute of Japanese Literature

日本語の歴史典籍の国際共同研究ネットワーク構築計画 (種別: 歴史的典籍研究事業)

English menu

《プレスリリース》「歴史的典籍オープンデータワークショップ～古典をつかって何ができるか！ じんもんそん2015～」を開催

●開催日時 ●プログラム ●対象 ●参加にあたってのご注意 ●お問い合わせ

日時 2015年12月18日(金) 15:00～18:00(開演15時)

場所 丸の内本館 6F 会議室 307
(東京都地下鉄丸の内線
法政大学前駅南口「丸の内線」下車、徒歩3分。)

参加費 無料

会場 人文文化研究棟 国文学研究資料館
1階 学術・文化研究棟 国文学研究資料館
後援 学術振興会 人文科学・コンピュータ研究会
日本デジタル・ニューメディア学会

使用言語 日本語

お申し込み 必ず申し込みサイトへお申し込みください
申し込みサイト 国立情報学研究所「情報学総合データベース」
①検索「『歴史的典籍オープンデータセット』」
②印刷「『歴史的典籍オープンデータセット』」
(http://www.nijl.ac.jp/cij/cijopen.html)

【開催趣旨】
国文学研究資料館では、歴史的なアイデアをもとにする視点による画像データ活用の可能性を
考え、進めています。国文学研究資料館の歴史的資料と画像データ約300点(全冊約3
万2千枚)を国立情報学研究所(NII)のご協力のもと、国文学研究所「情報学総合データリ
ポジトリ」より、データセットとして公開し、アイデアソンではこのオープンデータを活用し
ました。
国文学研究資料館・国文学研究所の両機関でアイデアを出し合い、おもむきとてい
し、参加型のイベント。

【プログラム】
①開会 歴史的典籍オープンデータとは
15:00～15:20 国文学研究資料館に関する大型プロジェクトの概要
15:20～15:30 オープンデータ101 (5分休憩)
15:30～15:45 質疑応答・アイスブレイク
(15:30～17:00 休憩)
②第1部 歴史的典籍アイデアソン
17:00～17:15 開演・エントランス
17:15～17:45 アイデアソン・セッション1
17:45～17:55 休憩
17:55～18:25 アイデアソン・セッション2
18:25～18:30 休憩
18:30～18:00 発表・まとめ
※プログラムは変更される可能性があります。

【対象】
研究者(人文系・歴史系に関わらず)、図書館員、学生の方、日本語の歴史典籍に関心のある
方。
※参加はグループ3～6名に限定して受け付けています。

国文学研究資料館 National Institute of Japanese Literature

日本語の歴史典籍の国際共同研究ネットワーク構築計画 (種別: 歴史的典籍研究事業)

English menu

《報告》「歴史的典籍オープンデータワークショップ～古典をつかって何ができるか！ じんもんそん2015～」グループ成果まとめ

国文学研究資料館は2015年12月18日に「歴史的典籍オープンデータワークショップ～古典をつかって何ができるか！ じんもんそん2015～」を開催しました。国文学資料の画像データと書誌データ約350点(全冊約6万4千コマ)を国立情報学研究所(NII)のご協力のもと、国文学研究所「情報学総合データリポジトリ」より、データセットとして公開し、アイデアソンではこのオープンデータを活用しました。このほど、アイデアソンで交わされたグループの意見をまとめましたので、ご報告いたします。このアイデアを実現できるよう、資料館では現在活動中です！

グループNo.6の発表

【江戸レシピの"つくれば"作成】

江戸時代の料理本に書かれているものを復元する試み。江戸時代の料理本は、正確な分量等を記載していないことも多い。だからこそ、その記述から、現代においていろいろな料理が出来上がる可能性もある。一般の人に江戸料理本をもとにした料理を実際に作ってもらい、そのレシピや作ってみた感想などのレポートをウェブにあげてもらおう。そこから、様々な料理の可能性を追求ができればと考えた。これは、古典の料理を元の通りに復元するのではなく、現代に活かす取り組みと言える。



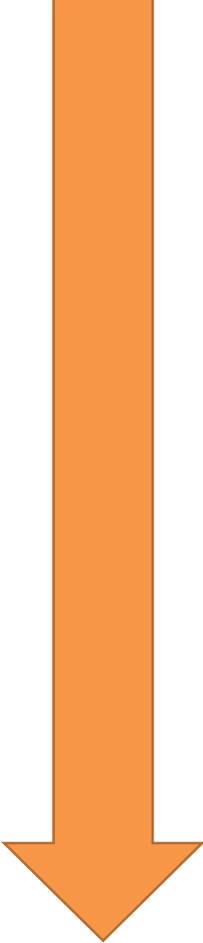
もちろん元となった古典の資料やその翻刻文、さらには現代語訳や挿絵などもウェブにあげ、一つの料理から多彩な情報や知識が広がる仕掛けを凝らし、SNS経由で「いいね！」を押してもらおうなどの工夫も必要。また、料理は地域ごとに違いがあるので(たとえばお雑煮の多彩さなど)、そうした情報をウェブにあげて比較できるようにすることも面白いと考えた。更に、古典を紹介した上で、日本料理の情報を発信することで、

海外の人へも日本文化の良さをPRすることが出来ると考えた。

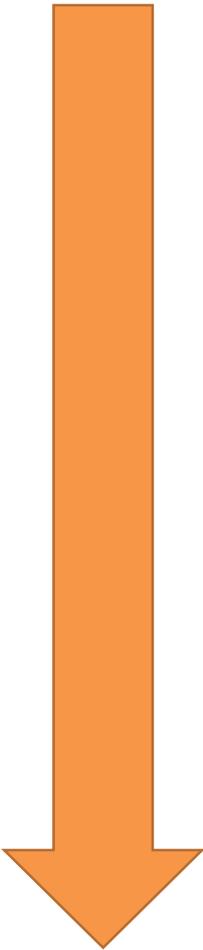
<http://www.nijl.ac.jp/pages/cijproject/ideathon.html>

http://www.nijl.ac.jp/pages/cijproject/report_20151218.html

アイデアの実現化

- 
1. 2015年12月 アイデアソン「じんもんそん」で料理本の存在を初めて知り、今回のアイデアを発案する。
 2. 2016年1月 クックパッドを訪問し、協力を依頼。
 3. 2016年11月24日（和食の日）江戸料理レシピデータセット公開 + クックパッドでもレシピ公開。
 4. 2017年6月 レシピを増やす。

江戸料理レシピデータセット

- 
1. 江戸の料理本をデジタル化
 2. くずし字を翻刻
 3. 翻刻を現代語訳
 4. 現代語訳をレシピ化・公開
 5. クックパッドでもレシピ公開
 6. つくれぽで個人の経験を共有

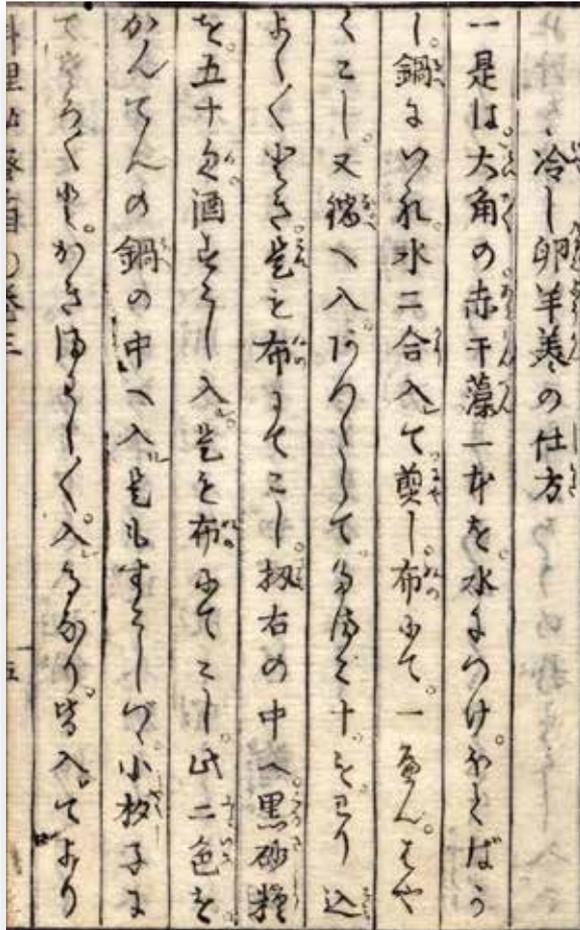
1. 江戸の料理本をデジタル化



日本古典籍データセット
(国文研所蔵)

- 『**万宝料理秘密箱** (まんぼうりょうりひみつばこ)』初版は1785 (天明5) 年。
- 「**卵百珍**」：素材としての卵一種で100の料理のバリエーションを楽しむ。
- 「**百珍物**」：食べるための食事から、食を楽しむ生活への変化という時代背景。

2. くずし字を翻刻

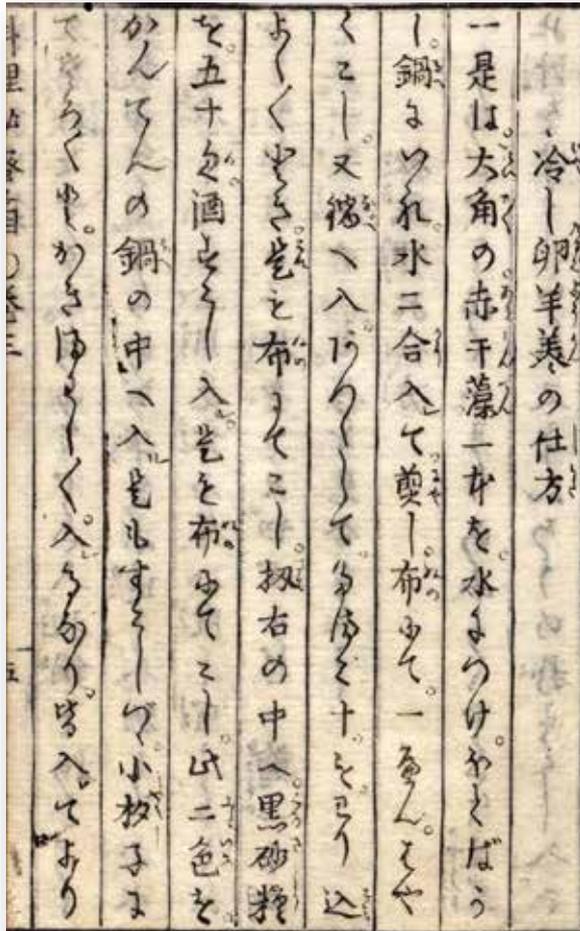


1	是は大角の赤干藻一本を水につけほとばかし
2	鍋にいれ水二合入して煎し布にて一へんはやくこし又鍋へ入れあつくして
3	たまご十ウをわり込よくよくとき是も布にてこし
4	扱右の中へ黒砂糖を五十匁酒すこし入ル是も布にてこし
5	此二色を かんてんの鍋の中へ入ル
6	是もすこしづつ小杓子にてそろそろとかきまわしかきまわし入れるなり
7	皆入してより又葛粉をすこし水にてとき入レ
8	扱鍋をぬき早く折敷にてもうちあげ平めに延し入レ物ともに水に入レ冷し遣ふ

日本古典籍データセット
(国文研所蔵)

江戸料理レシピデータセット (CODH制作)

3. 翻刻を現代語訳

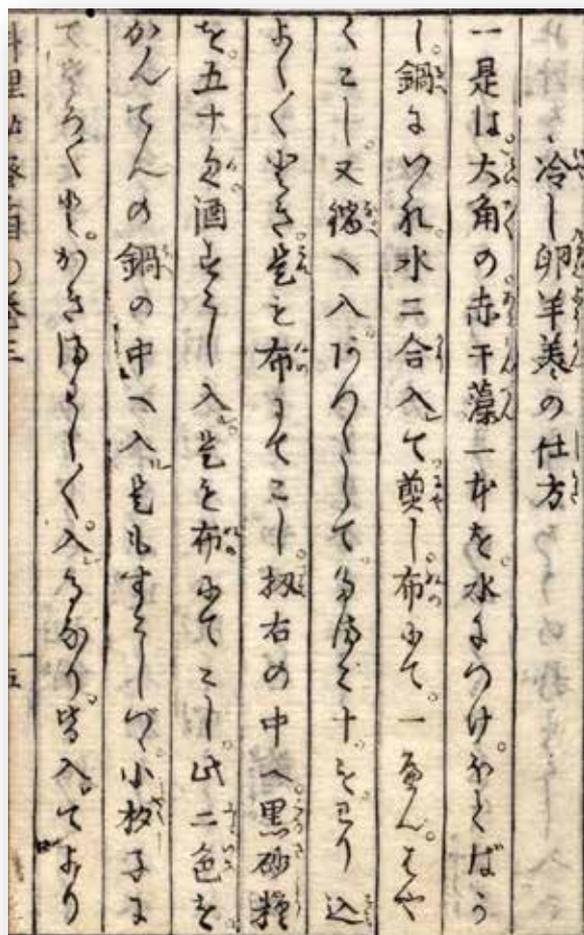


日本古典籍データセット
(国文研所蔵)

1	大きな赤寒天を1本水に付けてふやかす。
2	鍋に寒天と水2合 (360cc) を入れて煮溶かす。
3	を一度布で素早く漉し、再び鍋に入れて熱する。
4	生卵10個をよく溶き、布で漉す。
5	の中に黒砂糖50匁 (200g) と酒少しを入れ、布で漉す。
6	を寒天の鍋に入れる。小さな杓子で少しずつそろそろと混ぜながら入れる。
7	を全て鍋の中に入れたら、葛粉を水で溶き、鍋に入れる。
8	鍋を火から上げ、素早く中身を容器 (折敷) に広げ、平たく延ばし、容器ともに水で冷やす。

江戸料理レシピデータセット (CODH制作)

4. 現代語訳をレシピ化



日本古典籍データセット
(国文研所蔵)

1	寒天を水につけて、ふやかします。
2	生卵をよく溶きます。
3	溶いた生卵を布でこします。
4	黒砂糖と酒を入れ、溶かします。
5	4を3に入れ、再びこします。
6	鍋に寒天と水(180cc)を入れて煮とかします。
7	6を布などでこし、再び鍋に入れて熱します。
8	7の熱した寒天の中に、5の卵液を少しずつ入れます。
9	全て入れ終わったら、水でといた片栗粉を鍋に入れてさっと混ぜ合わせます。
10	鍋を火からあげ、中身を容器に入れます。
11	冷蔵庫で、2時間程度冷やします。

江戸料理レシピデータセット (CODH制作)

文字と写真による調理手順



江戸料理レシピデータセット (CODH制作)

オープンサイエンスをめざして - 社会と環境の情報学 -

現代の状況に合わせたレシピ

1. **材料の違い**：「葛粉」を「片栗粉」へと、現代でも入手しやすい材料に変える。
2. **道具の違い**：「容器ごと水で冷やす」を「冷蔵庫で2時間程度冷やす」へと、現代の道具に合わせて変える。
3. **分量の違い**：「卵10個」を「卵5個」へと現代の生活にあった分量に変える。

現代の文化を反映したレシピに「翻訳」

過去のことば・道具・文化

- **ト活卵**：蒲鉾を藁の太さ程に切る。
- **小豆餅卵**：ゆで卵を水飴の中に入れ、表面にくるりと塗る。
- **鶉卵**：鶉か鳩か　か雲雀か鴨か、その他どのような鳥にしても、鳥肉を丁寧によく叩く。
- **家主貞良卵**：上には、行灯の火皿を乗せ、この中に灰をうすく引き、火を入れて焼く。

江戸時代の文化を感じさせる記述から、過去の生活への想像を膨らませる。



料理研究家の写真

協力：グルメクッキングファクトリー



江戸料理レシピデータセット

日本古典籍データセットに含まれる江戸の料理本を、現代の生活にも取り入れるために、現代レシピに変換して提供します。

最初の江戸料理レシピとして、100種類以上の秘料理を集めた『万宝料理秘伝集 御百珍』を取り上げます。

「万宝料理秘伝集 御百珍」の江戸料理レシピ

くずし字を読める日本人が少ないという中で、日本古典籍データセットのようなデジタル画像を提供するだけでは、市民によるオープンデータ活用を進めることは難しいのが実情です。古典籍を日常生活にどのように活用していけばいいか、と考えているところで思い当たったのが江戸時代の料理本でした。これを現代語訳すれば現代でも料理を作って楽しめるのではないかと考えました。

雑煮などの季節の料理や地方色豊かな料理などは、日本人の生活に深く根ざしたものです。そして日本の料理としての和食は、単なる料理法を超えて自然の尊重という日本人の精神に基づく文化を表すとも言われています。平成25年には「和食：日本人の伝統的な食文化」がユネスコ無形文化遺産に登録され、和食文化に対する国際的な認知度も高まってきました。そんな和食という自身の文化をより深く理解するには、過去の料理について学び、気に向ければ作ってみることができるようなレシピデータが必要だと考えました。そこで以下のような「レシピ化」のプロセスに取り組みました。

データ概要

原本画像データ	日本古典籍データセットで公開する画像です。くずし字を読め、かつ江戸時代の日本語や料理法を知っていれば料理が作れます。
翻刻テキストデータ	原本画像のくずし字をテキスト化したデータです。江戸時代の日本語や料理法を知っていれば料理が作れます。
現代語訳データ	翻刻テキストデータの内容を現代の日本語に翻訳したデータです。江戸時代の料理法を知っていれば料理が作れます。
現代レシピデータ	現代語訳データの内容を、現代の道具や食材でも作れるものに実装し、食材の分量や写真を加えてより具体化したデータです。手順に従えば料理が作れます。

1. **翻刻**：全**107**点
2. **現代語訳**：翻刻107点中**40**点
3. **レシピ化**：現代語訳40点中**15**点（件数としては20点）

オープンデータ (CC BY-SA)
としてウェブサイト公開。

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

クックパッドでもレシピ公開

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」に参加。



レシピは写真つき、現代語訳は写真なし。元文献の記述やCODHページへのリンクも加えて公開。

クックパッド サービス一覧 お弁当におすすめのどんぶり料理 プレミアム

毎日の料理を楽しみに 253 レシピ

クックパッド江戸ご飯のレシピ

江戸時代のスイーツ 甘さスツキリ冷卵羊羹

江戸の料理本から見つけた和風スイーツです。プリン同様ですが、牛乳不使用でさらっとした菓糖の甘さがやみつきになります。

クックパッド江戸ご飯

材料 (2~4名分)	
卵	5個
寒天(赤)	1本(4g)
菓砂糖	100g
水	180cc
片栗粉	適量
酒	適量

カロリー:276kcal/人 糖質:0.3g/人

実器編 自動中 設定

1. 寒天を水に付けて、ふやかします。
2. 卵をよく混ぜます。
3. 溶いた卵液を布でこします。
4. 写真の如く、菓砂糖と酒を入れ、溶かします。
5. 4を3に入れ、再びこします。
6. 鍋に寒天と水(180cc)を入れて煮とがします。
7. 6を布などでこし、再び鍋に入れて熱します。
8. 7の熱した寒天の中に、5の卵液を少しずつ入れます。

<http://cookpad.com/recipe/4153357>

極めて大きな反応

人文学オープンデータ共同利用センターさんがリツイート

うずら @caille2006 · 11月26日
このプロジェクトがすごいのは、古文書の情報をさらに現代の生きた情報にするために、クックパッドにアカウントを開設してレシピを公開し「つくれば」も受け付けていること。江戸ご飯とつくればというこの未来感パネい。 cookpad.com/kitchen/146046...



クックパッド江戸ご飯 のキッチン

プロフィール

トップ	レシピ 32	つくれば 0	献立 0
-----	-----------	-----------	---------

レシピを検索

7317 リツイート

<https://twitter.com/caille2006/status/802575840819089409>

2018/2/23

人文学オープンデータ共同利用センターさんがリツイート

NII 国立情報学研究所(NII) @jouhouken · 11月24日
[プレスリリース]
江戸の文化を現代に取り込む「江戸料理レシピデータセット」を整備～江戸時代の料理本を「レシピ化」し、クックパッドでも公開～
nii.ac.jp/news/2016/1124



← 1 1,074 971

1052 リツイート

<https://twitter.com/jouhouken/status/801693251052781568>

人々の反応の違い



私：「こんど江戸料理のレシピをウェブサイトで公開するんだ。」
相手：「ふーん...」
（関心なし）



私：「江戸料理のレシピはクックパッドでも公開するんだ。」
相手：「それはいいね！」
（顔が輝く）

なぜ人々の顔は輝くのか？

1. クックパッドと聞くだけで、**データの
利用方法**が自動的に思い浮かぶ。
2. クックパッド形式にデータが整備されると、**料理の再現性**に確信が持てる。
3. 馴染みのプラットフォームに入ると、
「**自分たちのデータ**」と感じる。
4. クックパッドという現代と江戸料理と
いう過去のギャップに意外感がある。

古いものと新しいもの

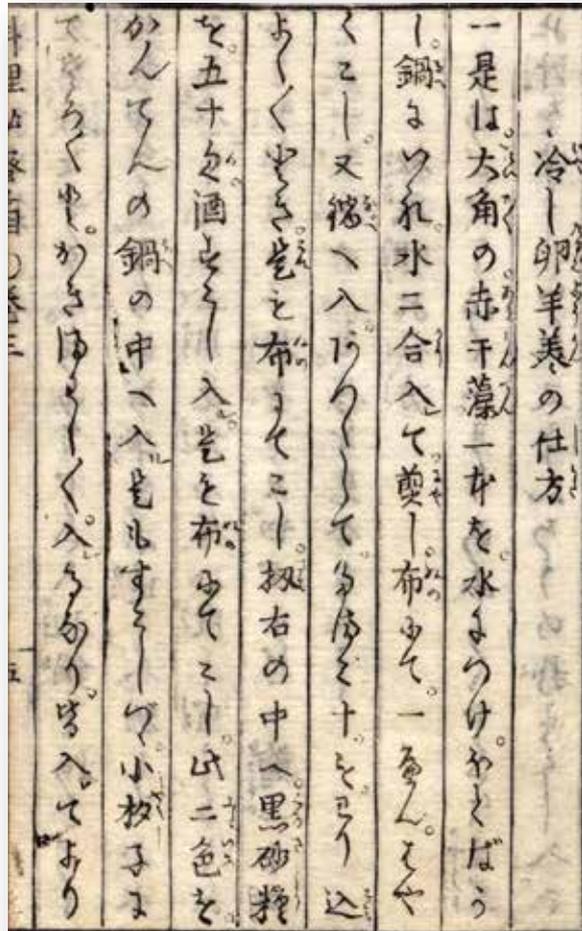
聖書「新しいぶどう酒を新しい革袋に入れる」
→「古いものに新しいものを継ぎはぎして、それでうまくやっていると考えているが、それによって古いものをも新しいものをも駄目にしてしまうようなことをしてはいないのか」



古いものに新しいものを継ぎはぎして、うまくやろうとしていないか。これはオープンサイエンスにも通じる問いかけではないか。

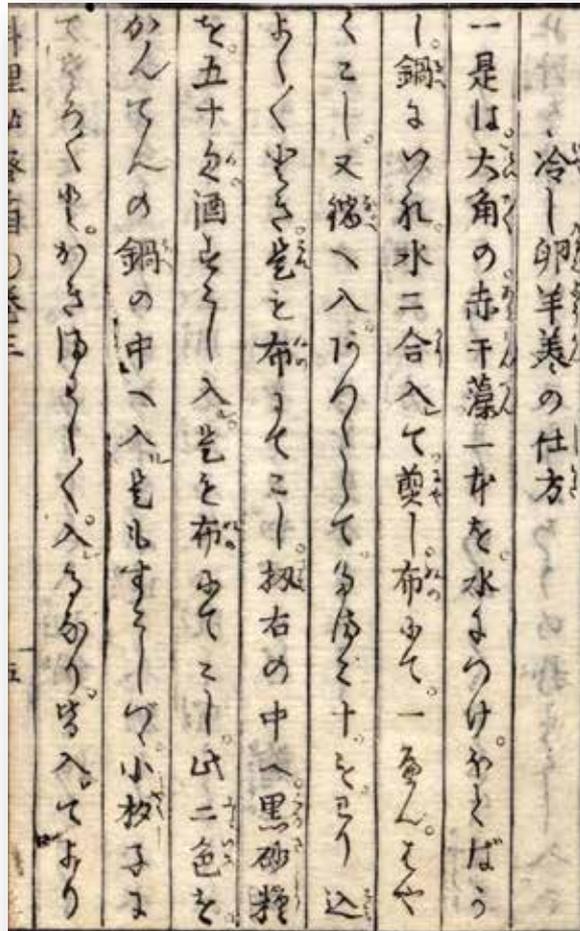
参考：日本キリスト教会 南柏教会、新しいぶどう酒は、新しい革袋に、
<http://minami-kashiwa.sakura.ne.jp/2015/08/02/704>

古いコンテンツを 古いやり方で



- 書籍をデジタル化しても、**アトムをビットに変換**したに過ぎない。
- 情報の形式は昔のまま。たとえ翻刻しても、この点は改善しない。
- **古いもの（本）に新しいもの（ウェブ）を継ぎはぎ**したただけではないか？

古いコンテンツを 新しいやり方で



どのプラットフォームを選ぶべきか？

新しいプラットフォームの仕様に合わせて、どのように情報形式を変換するか？



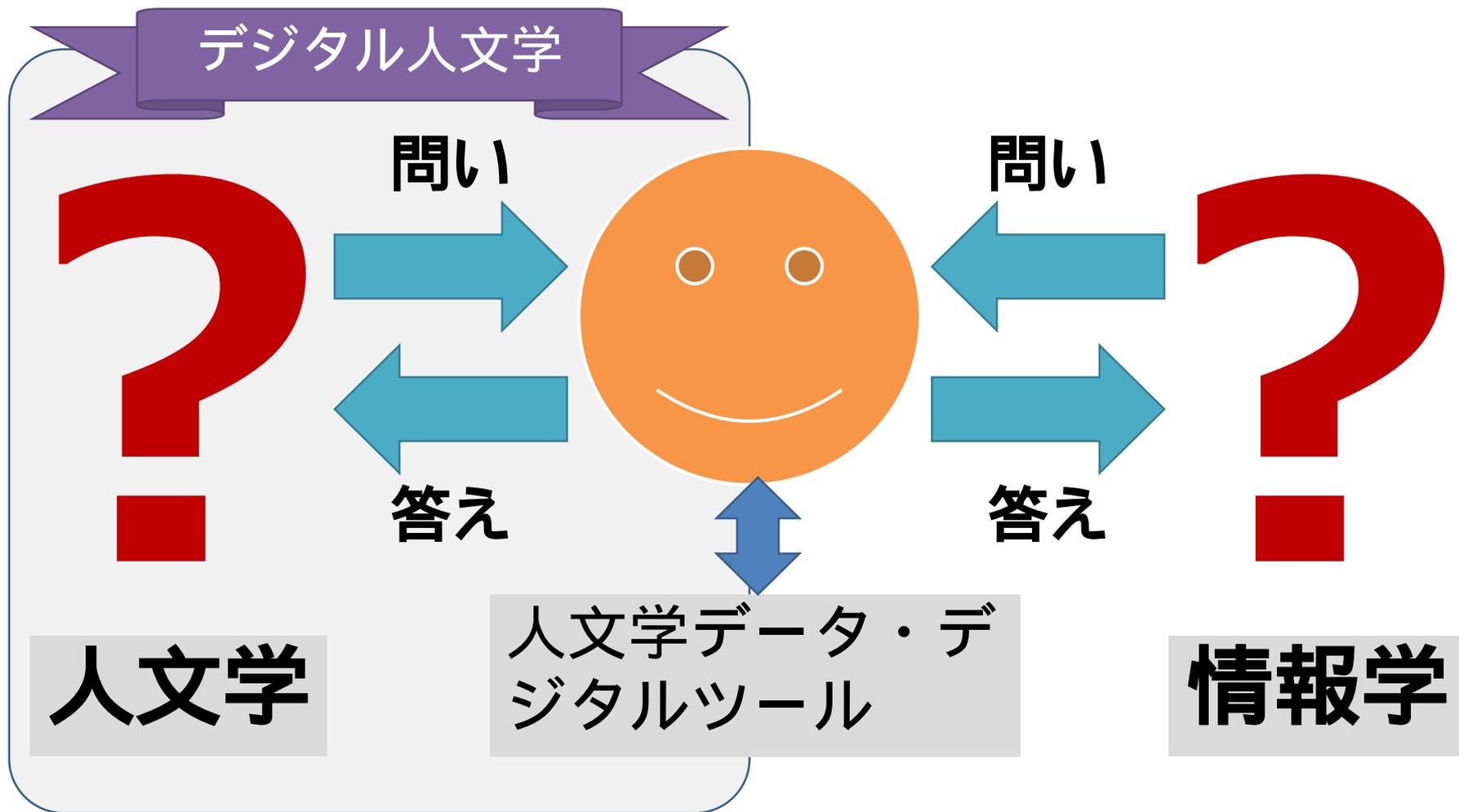
クックパッドへの「展示」



- たまたま手元にあったデータをアップロードしたのではない！
- 書籍の選定、料理の選定、手順記述、テキスト表現、写真撮影など。
- 最初からクックパッドへの掲載を想定した方針決定を行っている。

おわりに

学際領域と同床異夢



情報学の発想

1. **抽象化**：実世界をモデル化し、単純化して、操作可能な体系とする。
2. **手順化**：手順を実際のシステムに落とし込みながら、対象を深く理解する。
3. **概念化**：研究の対象を言語化し、目指すべき方向やゴールを明確化する。
4. **一般化**：同様の抽象化・システム化・概念化などを、類似対象にも展開する。

手段と目的の交換

1. 人文学における手段は、情報学における目的になりうる。
2. 情報学における手段は、人文学における目的になりうる。
3. 学問Aと学問Bが相補的な関係になれば、お互いにハッピーになれる？
4. そのためには、**相手の価値観を理解し、尊重する**必要がある。

人々が本当に欲しいもの

People don't want to buy a quarter-inch drill. They want a quarter-inch hole!

(人々が欲しいのは、ドリルではなく穴である) - Theodore Levitt

ドリル = 手段 **穴 = 目的**。目的を実現するには様々な手段があるが、特定の手段にとらわれると、最適な手段が選べない。

持続可能性と多様性

1. 一つの分野で成功するには、特定の価値観に染まったほうが有利。
2. 一つの価値観に染まりすぎると、越境することは難しくなる。
3. 越境できるオープンサイエンス人材の育成には、評価軸の多様性が必要。
4. 研究生態系の多様性がオープンサイエンスの持続可能性につながる。

お知らせ：CODHセミナー

- **【第6回CODHセミナー】歴史ビッグデータ
～過去の記録の統合解析に向けた古文書データ
化の挑戦～**
- **古文書に埋もれた記録を拾い集めることで、過
去の世界の統合解析は進展するか？**
- 日時 2018年3月12日（月） 13:30-18:00
- 会場 国立情報学研究所 1208 / 1210会議室
- 参加申し込み受付中です。
- <http://codh.rois.ac.jp/seminar/historical-big-data-20180312/>

謝辞・参考資料

- NIJL-NWプロジェクトでは、国文学研究資料館 古典籍共同研究事業センターの協力を得ました。
- PRMUアルゴリズムコンテストを開催いただいた、PRMU研究会担当の皆様には感謝します。
- 人文学オープンデータ共同利用センター
 - <http://codh.rois.ac.jp/>
- オープンサイエンス
 - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>