

情報学と人文学の出会い ～ デジタル・シルクロードから 人文学オープンデータ共同利 用センターへ～

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/>

“Meets”（出会い系）の流行



Machine Learning Meets Economics

12th Mar '14, 01:28 PM in Video

MIT Technology Review

Topics+ The Download

Discover the emerging technologies that will **CHANGE OUR WORLD**

View from the Marketplace

Customer Data Meets AI

A new day is dawning for the customer experience, driven by the application of artificial intelligence, machine learning, and automated technologies to CRM data. The potential exists to transform the customer's experience by providing service in a more predictive and intuitive way than ever before.

by MIT Technology Review Custom July 17, 2017

IBM

IBM Research

AI meets IoT

feedback



Deep Genomics: Artificial Intelligence Meets the Human Genome

by VLAB

KDD2017 Workshop

Machine learning meets fashion

Data, algorithms and analytics for the fashion industry
14 August 2017, Halifax, Nova Scotia - Canada





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。

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

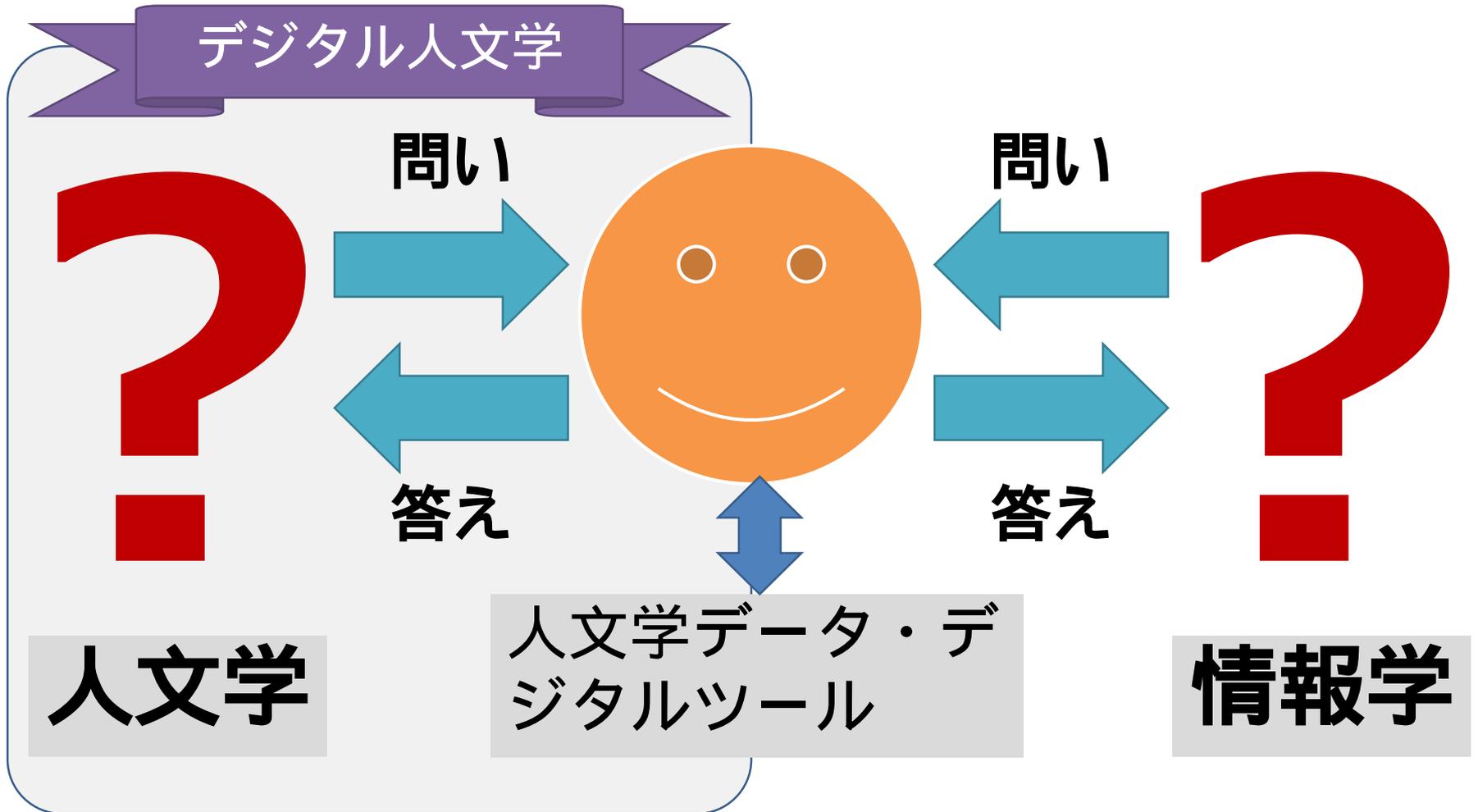
情報学と人文学の出会い

1. 人文学データに触れたきっかけは、プロジェクトに参加させられたから。
2. ピュアな理系（？）ではないので、人文学データの中身にも興味を持てた。
3. デジタル・ヒューマニティーズは、私から見れば人文学のカルチャーが強い。
4. **情報学の発想を大切にしつつ、人文学としての価値も意識する。**

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。

学際領域と同床異夢

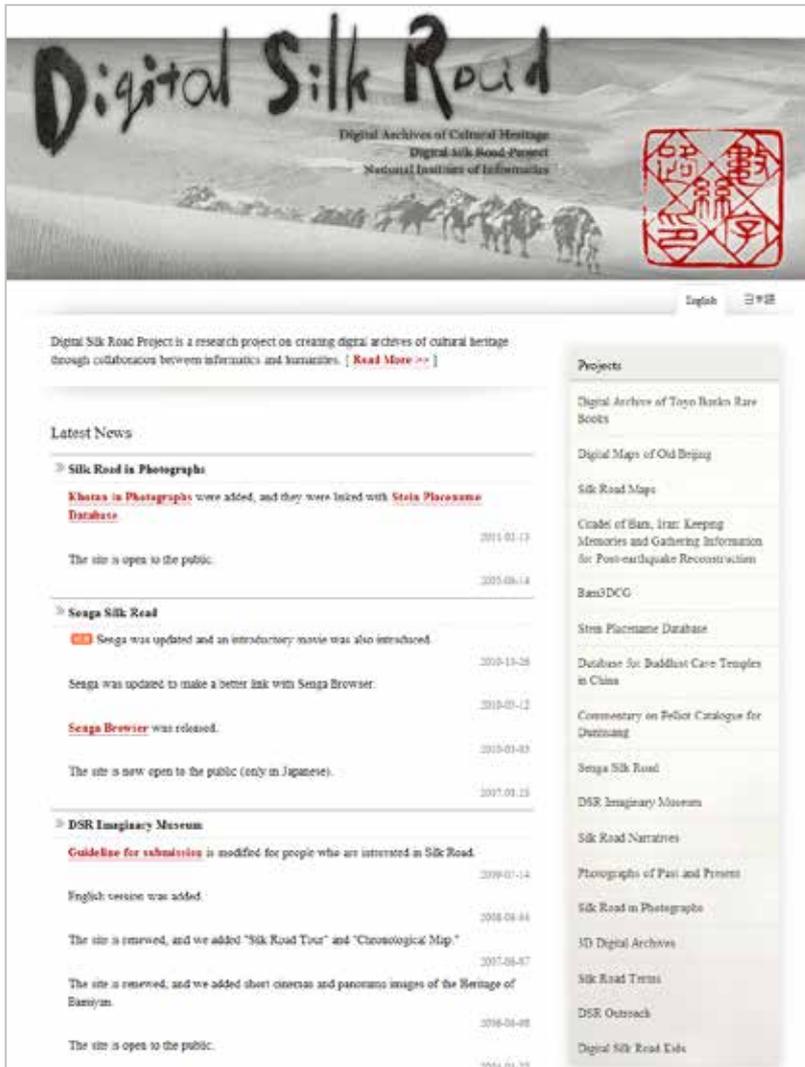


デジタル・シルク ロード

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。



東京 / 奈良シンポジウム

ENGLISH

開催概要
概要
プログラム掲載
参加募集
申込案内
関連経路
申込募集
9/19日まで締め切りました
Contact Information
Contact Us

Call for Participation
「デジタル・シルクロード」東京シンポジウム
Tokyo Symposium for Digital Silk Roads



主催：ユネスコ国際教育科学文化機関、日本ユネスコ国内委員会、国立情報学研究所
後援：総務省、外務省、経済産業省、(社)日本ユネスコ協会連盟、(財)デジタルコンテンツ協会、デジタルアーカイブ推進協議会、アドバンスドメディアフォーラム/研究会
協力：中華人民共和国北京政壇文化院/院團
日程：平成13年/2001年12月11日(火)～13日(木)
会場：学術総合センター(東京都千代田区一ツ橋2-1-2)

Subj:
「デジタル・シルクロード」東京シンポジウムは、文化とデジタル技術の融合により、新機軸でのシルクロードについて、発表と討論の場を提供し、数千年の昔から東西世界を結ぶ、文化遺産の帯であるシルクロードに、デジタルが加わり、未来に継承をへた多くの文化遺産が保存、活用、長年にわたる人類の歴史を明らかにするなどの役割が期待されています。本シンポジウムによる文化遺産保存や研究の促進が期待されています。また、海外・国内・産学・産官学連携によるデジタル技術の活用や、各国に存在する研究資料を世界で共有し活用して、公開することにより、より一層の文化遺産保存への貢献が期待され、国際的な共同研究の促進が期待されています。また、デジタル技術分野の技術者にも向け、さらに一歩を踏み出すためのシンポジウムの目的です。

●12月11日のセッションは公開セッションとして、一般の参加希望者にも広く公開する形で開催します。参加費は無料です。●



一般公開セッションでの講演



カーン：ユネスコ事務局長補



白川哲久 文部科学省国際統括官
(日本ユネスコ国内委員会事務総長)



学術総合センター1階ロビーでの展示



専門セッションでのパネル討議



末松安晴 国立情報学研究所長

開会式での各主催者からの挨拶

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

http://dsr.nii.ac.jp/tokyo_symposium_nii_news.pdf

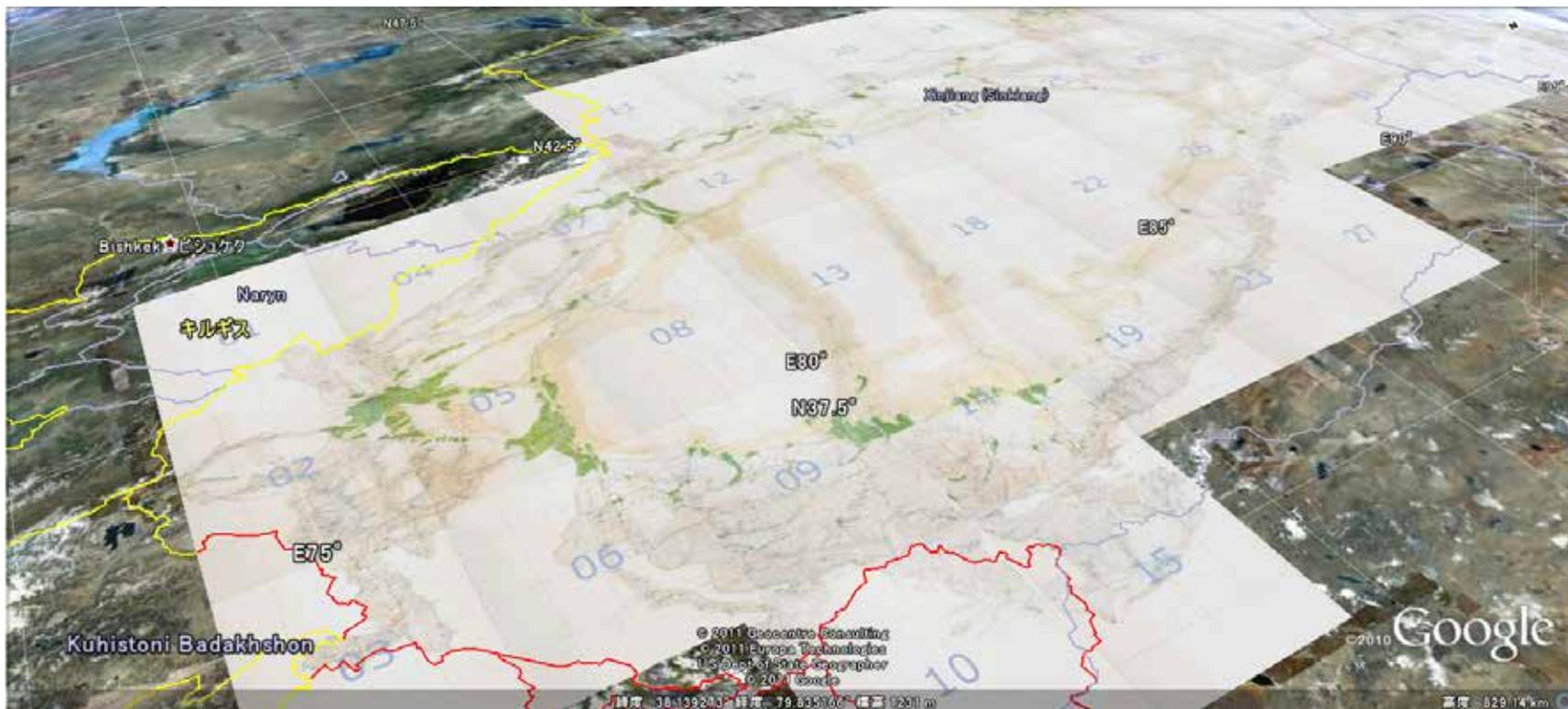
1. 東京シンポ（2001年12月）は参加者330名。
2. 奈良シンポ（2003年12月）は参加者240名。



東洋文庫貴重書 アーカイブ

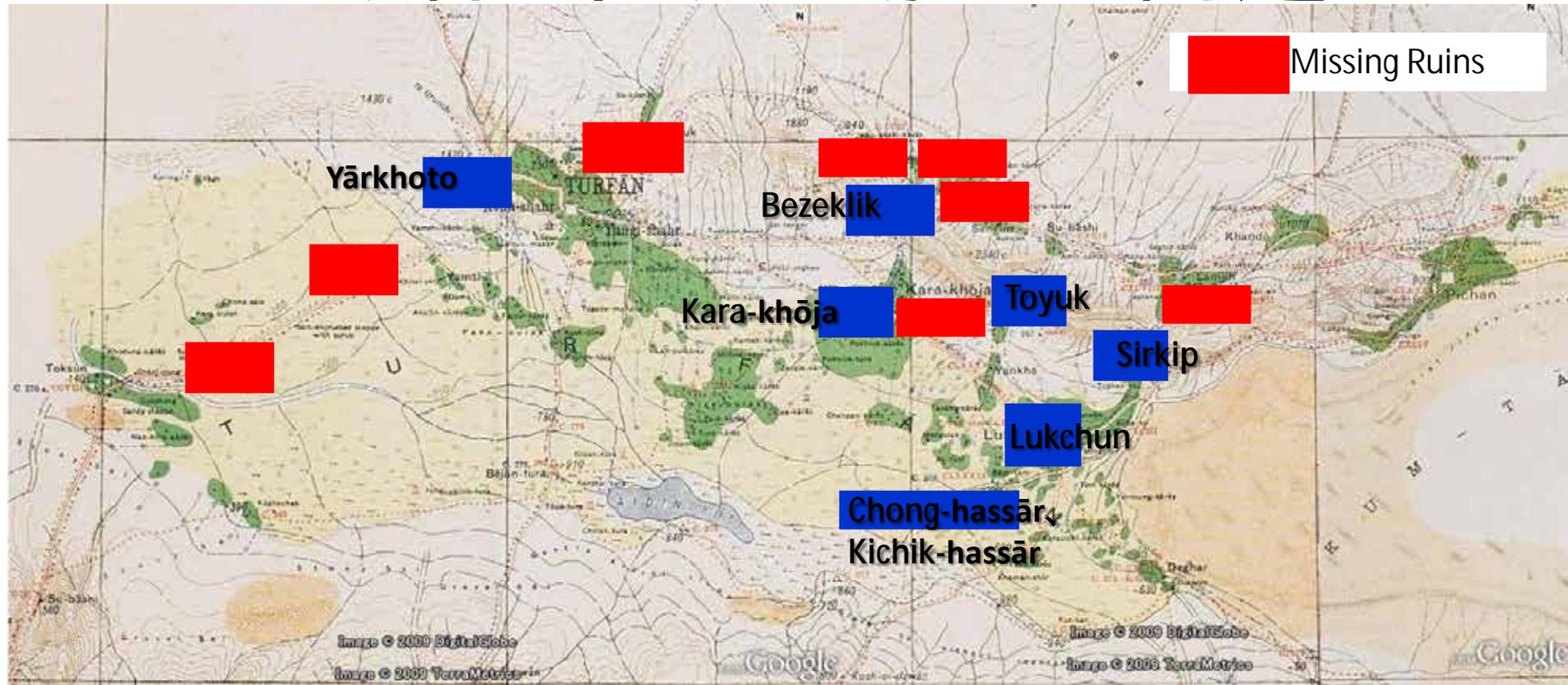
- 245冊の書籍72,591ページをデジタル化・公開。
- 研究コミュニティ内で最重要な書籍を選定。
- キャプションや目次などは人手で入力。
- 全文テキストはOCR活用（誤りは未訂正）。

スタイン地図 (シルクロード)



信頼されていた地図に、実は誤差があった？

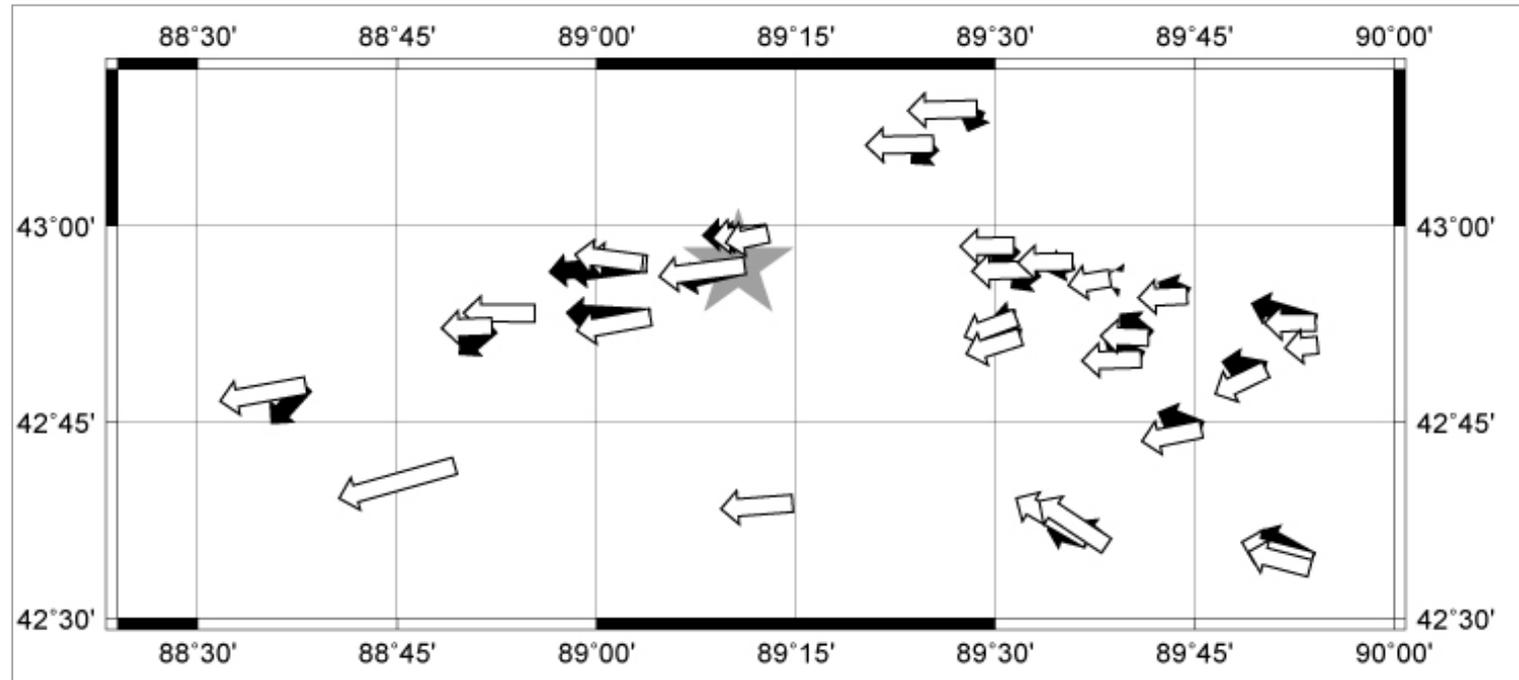
「所在不明遺跡」の問題



Oi-tam, ruined fort
Bögan-tura
Buluyuk (Shipang, Sassik-bulak, Kazma)
Murtuk-ruins

Yoghan-tura
Chikkan-köl
Bedaulat's town, Bēsh-kāwuk, Kosh-gumbaz
Yutōgh

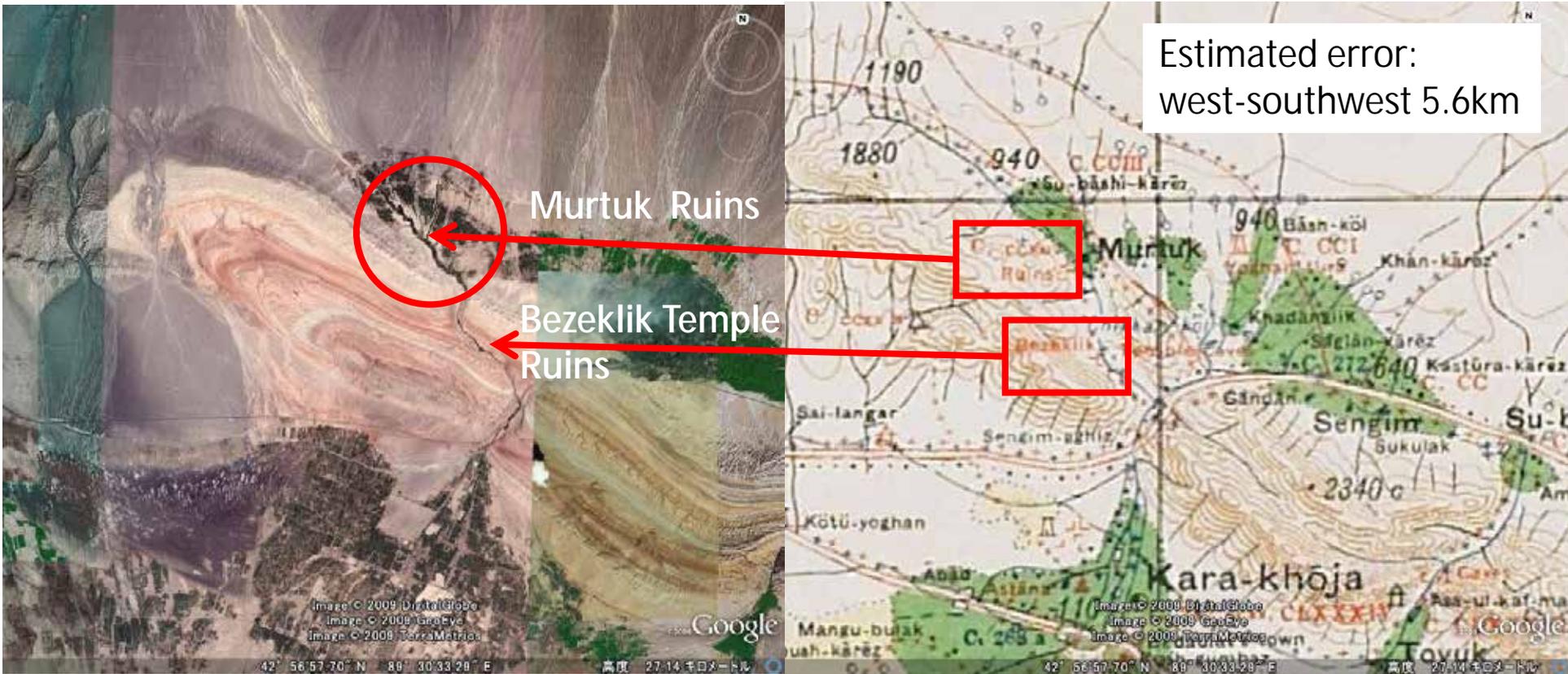
トルファンの古地図誤差分布



Error Distribution in Turfan Basin / White: Innermost Asia / Black: Serindia

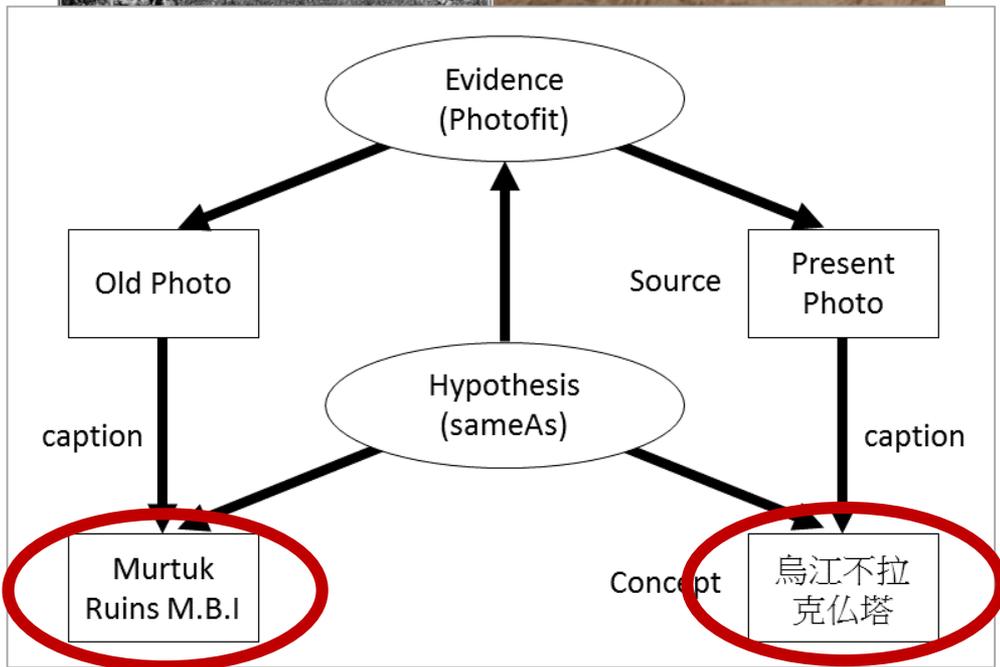
経度方向に誤差が大きいのは、20世紀初頭
は経度測定が技術的に難しかったため。

ムルトウク遺跡の同定



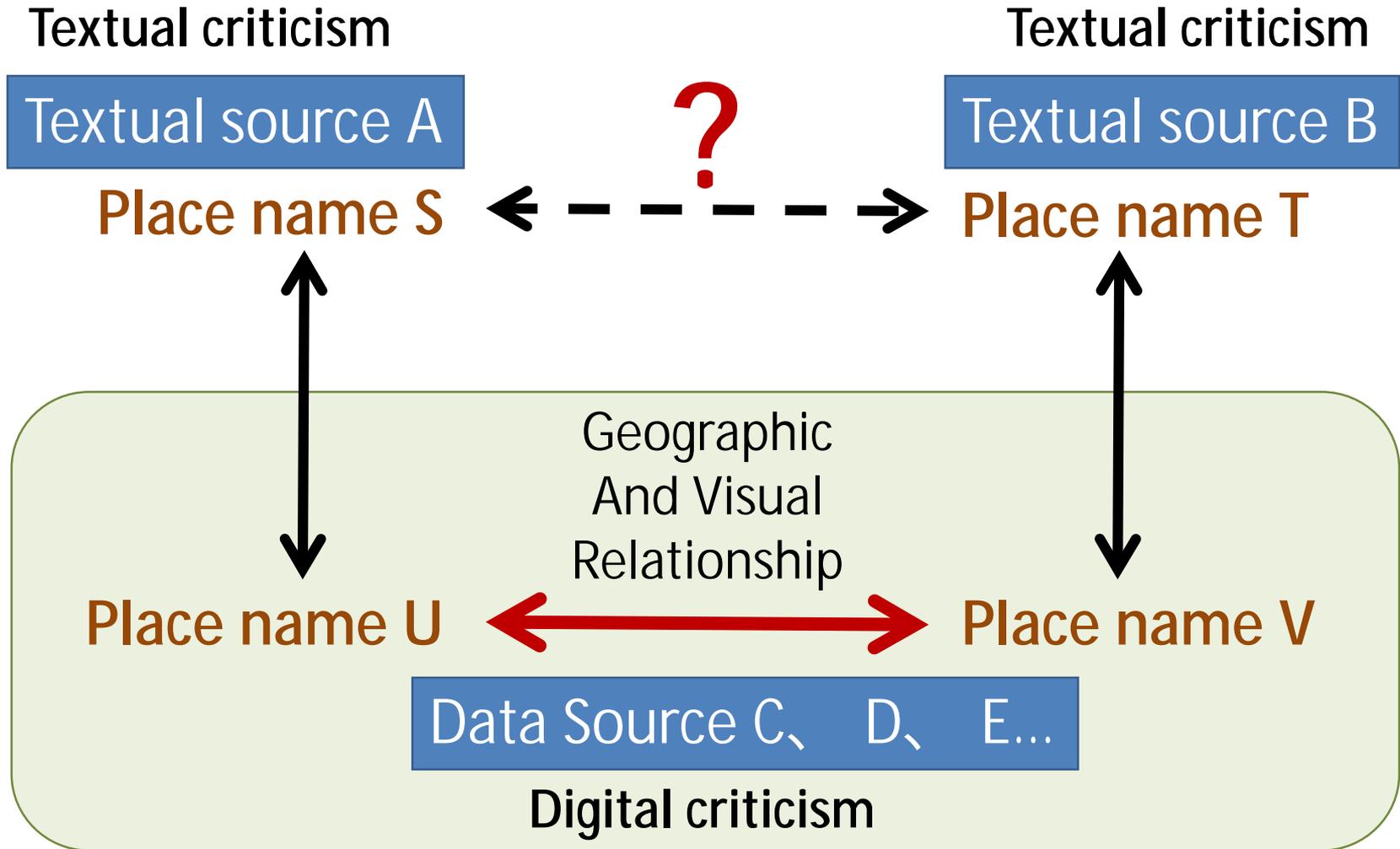
周辺の誤差推定に基づき、ムルトウク遺跡の位置は赤丸の位置と推定できる。

Murtuk Ruinsと烏江不拉克仏塔



ある古写真中の「Murtuk Ruins」という名前の遺跡と、ある現代写真中の「烏江不拉克仏塔」という遺跡は、同じものなのか？

地名のマッチング



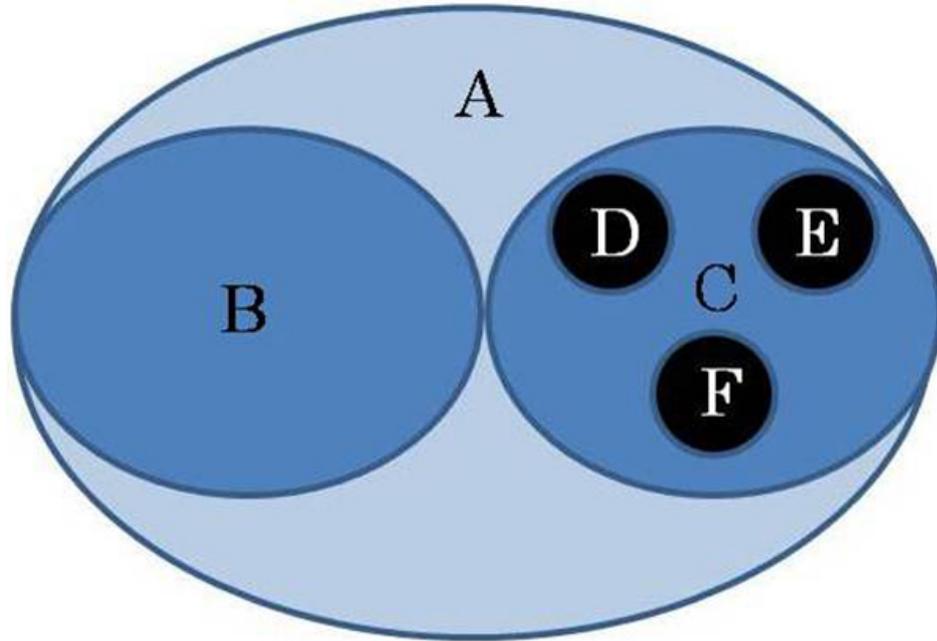
遺跡のマッピングと統合



<http://dsr.nii.ac.jp/digital-maps/stein/place-names/>

スタイン地名データベース：タリム盆地周辺の地名6,303件をデータベース化。現代の遺跡名や現代の写真とも統合を進めている。

遺跡データベースの構築



遺跡の統合：ある史料の遺跡Aは、別の史料の遺跡BやCを含み、それは別の史料の遺跡D・E・Fを含む。

- 100年前の探検隊が残した情報と、現代の発掘から得られる情報とを「**リンク**」。
- **複数の史料からより多くのエビデンス**を集め、歴史像を描くことができる。

デジタル史料批判

Digitally-Enabled Criticism

- 歴史学の研究で、**地図や写真などの空間画像史料を批判的に解釈するには、デジタルツールの開発が不可欠である。**
- 空間画像史料の適切な解釈は、**歴史学における新しい発見につながる。**
- より詳しくは：西村 陽子, 北本 朝展, "デジタル史料批判と歴史学における新発見", 人工知能学会誌, Vol. 31, No. 6, pp. 769-774, 2016年11月

人文学オープンデータ 共同利用センター

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/>

研究者のためのオープンデータ

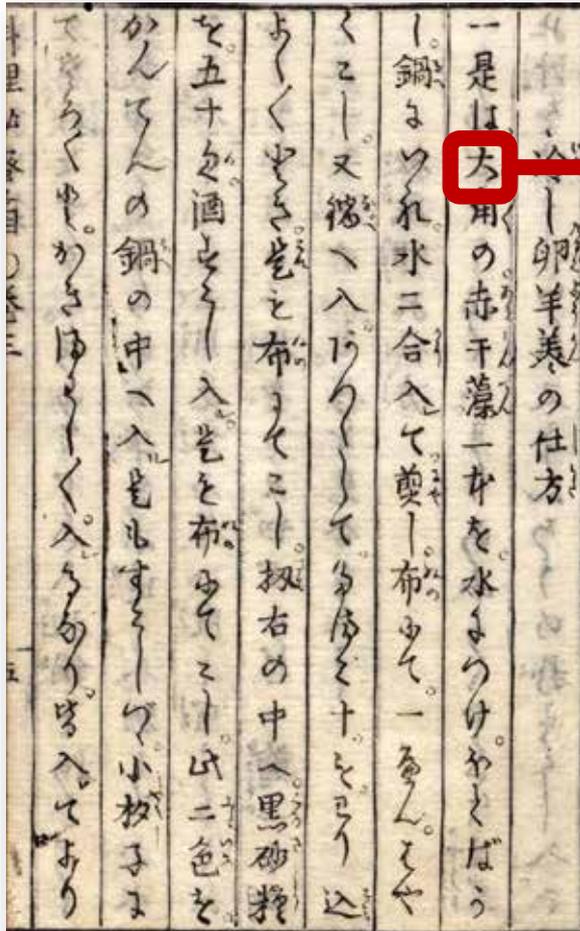
<http://codh.rois.ac.jp/pmjt/>



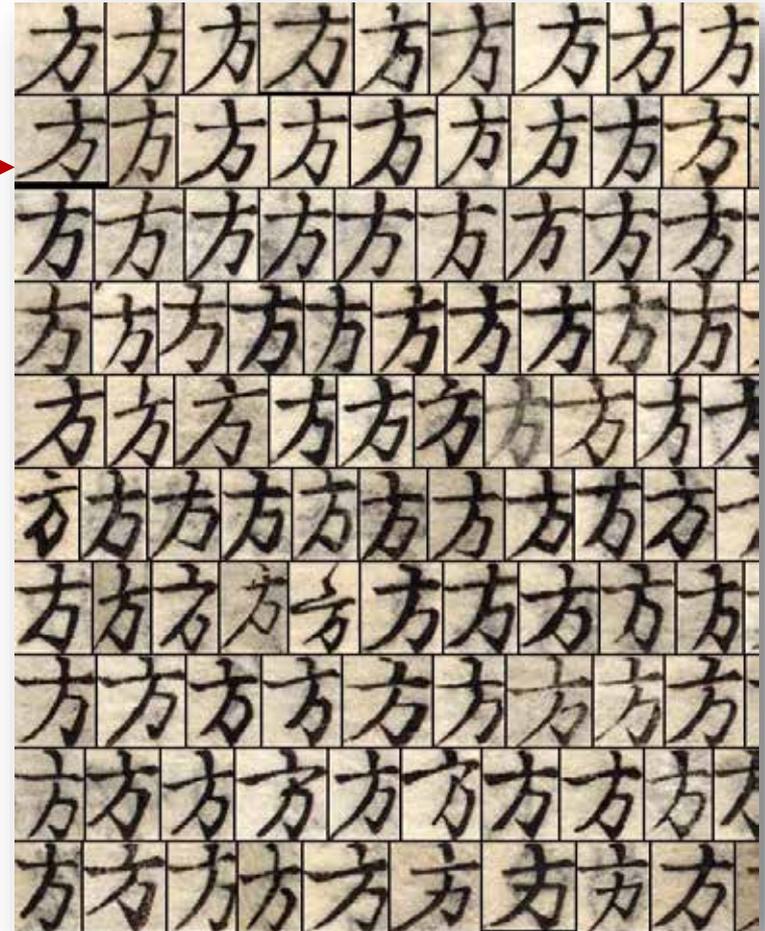
日本古典籍データセット（国文研所蔵）

機械のためのオープンデータ

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



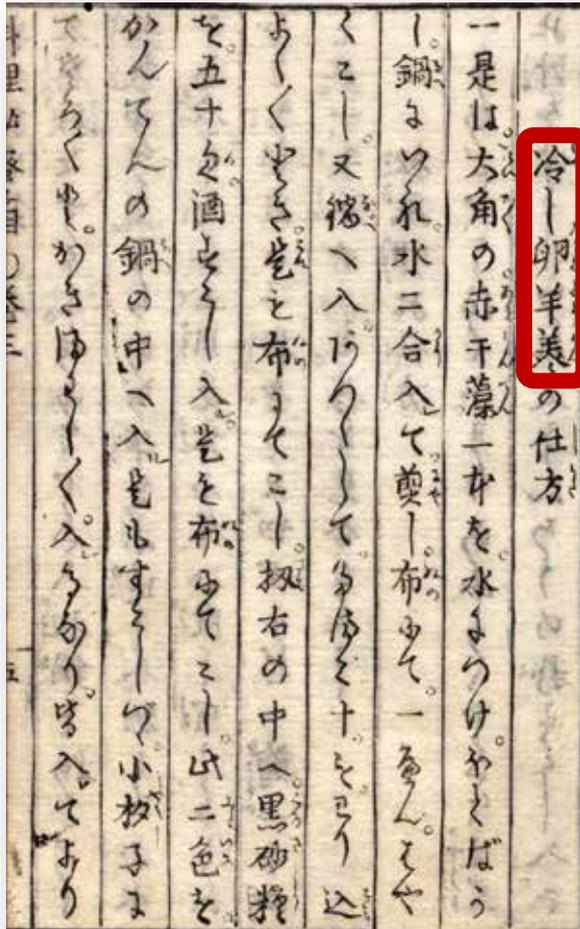
日本古典籍データセット
(国文研所蔵)



日本古典籍字形データセット
(国文研所蔵・CODH加工)

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

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。

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

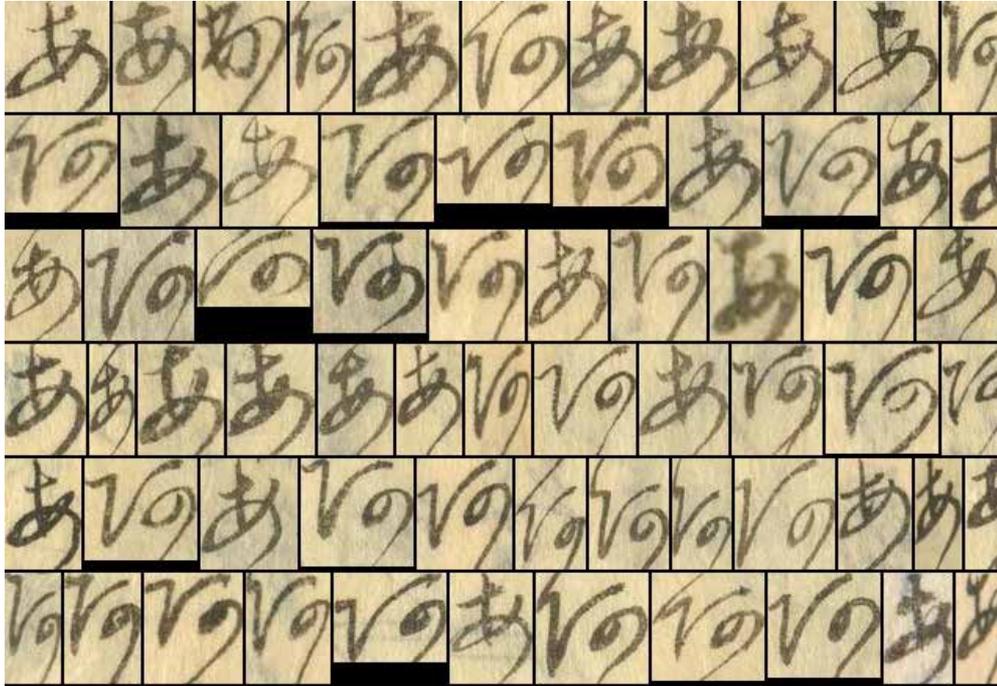
原本補正画像データ	日本古典籍データセットで公開する画像に対して、翻刻作業を容易にするための前処理として、見開き画像を分割するとともに、回転させて正立させるという処理を加えた画像です。
文字座標データ	原本補正画像データ上で、文字を取り囲む長方形の座標 (X,Y,W,H)、文字のUnicodeコードポイント、ブロックID、文字IDを記録したものです。
字形画像データ	「原本補正画像データ」に「文字座標データ」を適用して切り抜いた画像であり、文字種ごとに字形を閲覧しやすするために提供するものです。
作業報告文言	作業で成りなかつた文字に関する情報や、その他の注意事項を記したドキュメントです。

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。

人間のための学習データ



変体仮名「あ」

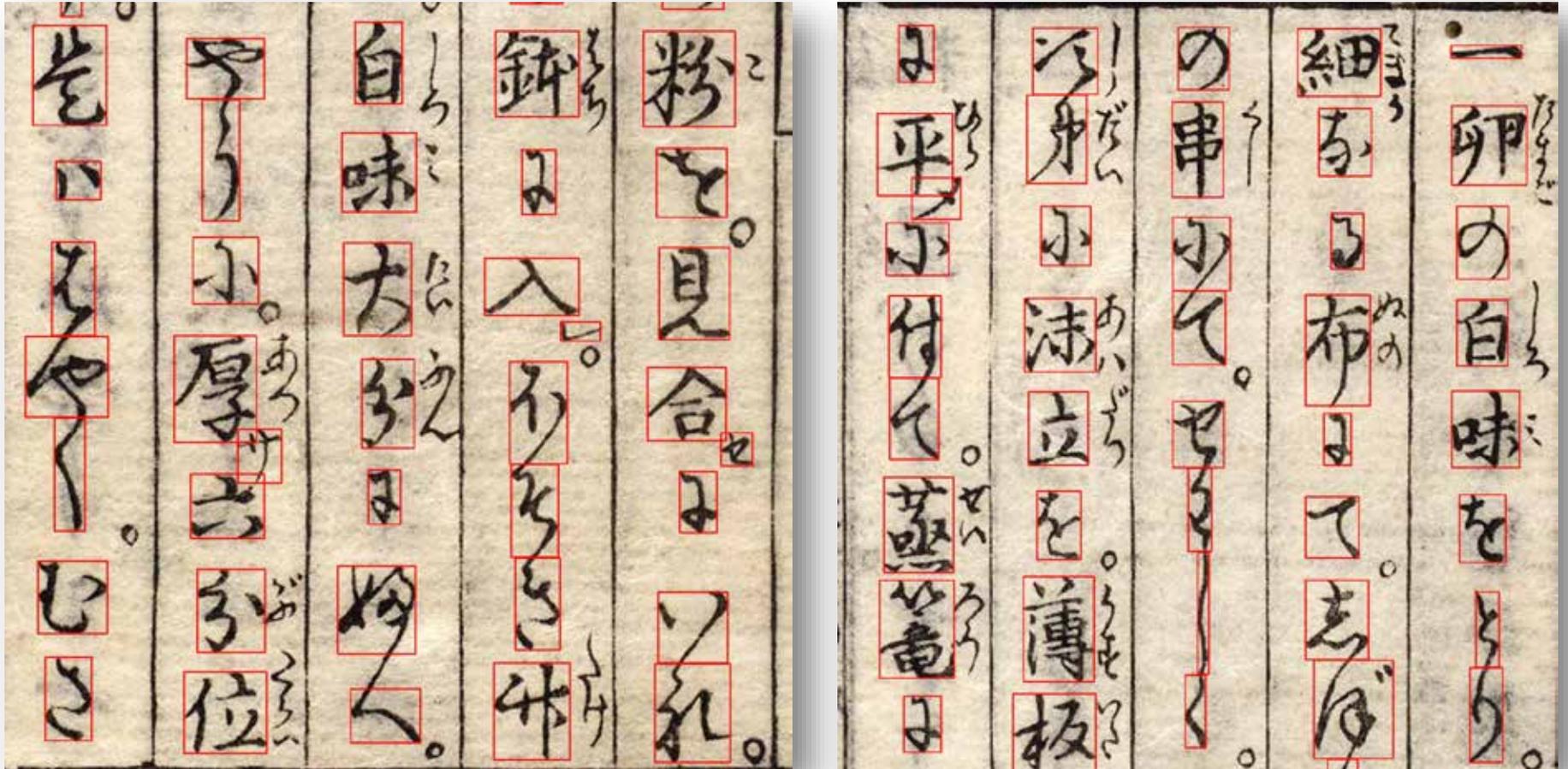
- 字形を目で見て確認できる。
- **学習アプリの素材にもなる。**
- くずし字が読める人が増える→データの利活用が進む→くずし字がより広まる。

機械のための学習データ

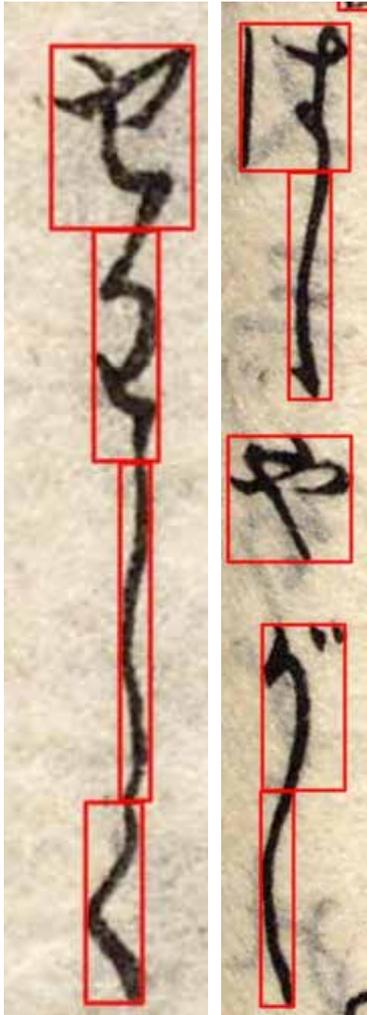
文字種	文字数
に	15,982
の	14,337
し	13,386
て	12,976
と	11,068
を	9,208
は	8,725
り	8,656
か	8,439
な	7,892
3,999文字種	403,242文字

- 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。

字形画像データ



分割と認識の深い関係

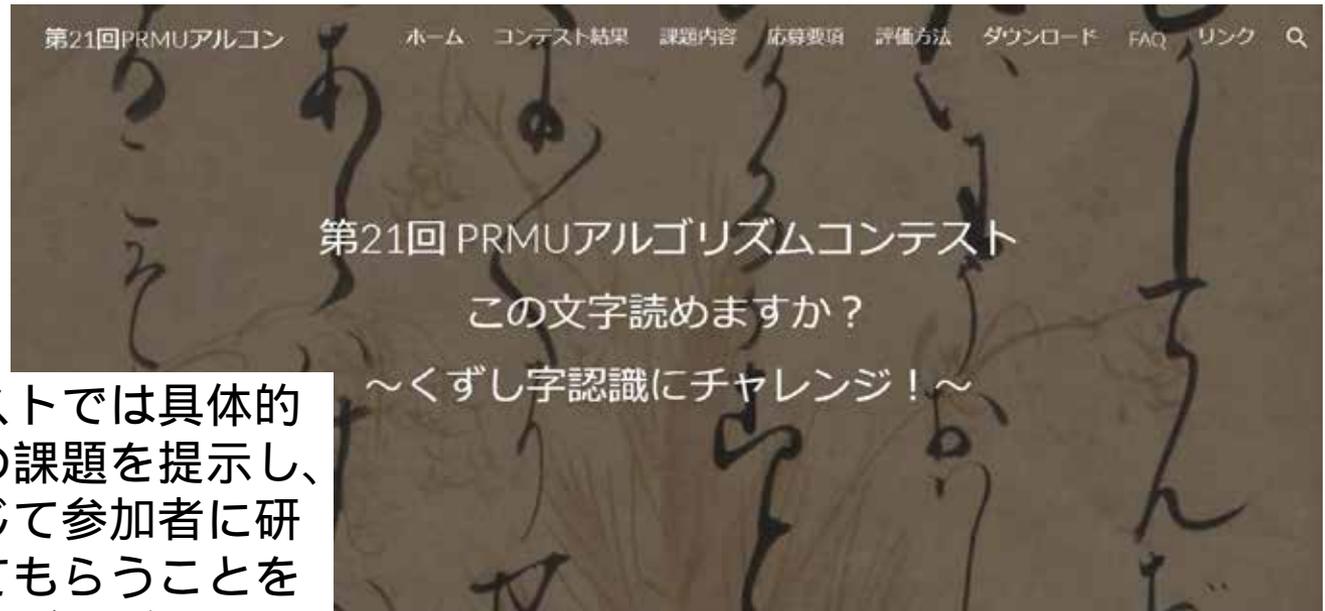


1. 文字を認識するには、文字を分割しなければならない。
2. 文字を分割するには、文字を認識できていなければならない。
3. 分割の自動化が必要かどうかで、問題の難易度は大きく異なる。
4. 分割という難問を回避する文字認識手法も研究が進んでいる（はこだて未来大学の寺沢氏による）

AIはくずし字を読めるか？

- **ディープラーニング**：技術の進展に伴って、難問が解ける可能性が出てきた。
- **ディープアクセス**：過去の日本文化の詳細に対する情報アクセスを実現する。
- **くずし字チャレンジ**：文字認識に加え、あらゆる技術や知識を総動員する。
- 遺伝子を全解読する**ゲノム解析** 古典籍を全解読する**スク립トーム解析**

PRMUアルゴリズムコンテスト



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>

課題内容



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。

マッチングと比較

過去と現代の景観比較

今昔写真（2007年）

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



1. ノートPCに古写真画像を表示。
2. カメラのファインダー内の景観とPCの古写真画像を比較。
3. 脳内でシミュレートし、最も一致する時点でシャッターを押す。

視線が動くため撮影が難しい！

北京古景観（2008年）

<http://dsr.nii.ac.jp/beijing-maps/>



1. Google Earth上に「乾隆京城全図」古地図を表示。
2. 古地図と古写真の撮影角度を合わせて、撮影位置を設定。
3. GEの3次元空間中に同時表示。

視点が異なるため見づらい！

メモリーグラフ

記憶を重ねる新しい写真術 (Memory+graph)



「メモリーグラフ」アプリ



<https://play.google.com/store/apps/details?id=jp.ac.nii.dsr.cias.memorygraph>

【謝辞】カメラにはOpen Cameraを利用。こうしたライブラリがないと、多機種をカバーするのは不可能。

- 2016年11月5日にGoogle Play公開。
- 一部の機能は準備中。近日中にリリースしたい。
- iOS版は、Android版が成熟した後？

同一地点・同一構図の写真



出典：「大好き神田」歴史写真館「昭和19年1月戸田武雄氏の出征」<http://www.daisuki-kanda.com/>
提供：小川町三丁目西町会創立50周年記念誌

写真のマッチング

拡張現実 (AR) ← → アクティブファインダー

古写真A

- 実空間に古写真がマップされる。
- **展示**のためのツール。

景観A

可変ファインダーは、人文学的資料を対象とした時間変化を探索するフィールドワークツール。この目的はARとは異なる。

古写真B

- 古写真に実空間がマップされる。
- **参加**のためのツール。

景観B

東京古写真ハンティング



- 2015年2月21日：国際オープンデータデーの日に開催。
- グループに分かれ、古写真の撮影場所を議論しつつ探索。
- TileJSONを使って古地図表示。「歴史的農業環境閲覧システム」「今昔マップ」を利用。

ベスト・マッチングの追求



三宅坂より司法省を望む、写真の中の明治・大正、東京編

- これだけ風景が明確な写真であっても、撮影地点を探すのは簡単ではない！
- 最初一枚の撮影に30分以上（しかも不正解！）。



正解を得るまでの試行錯誤



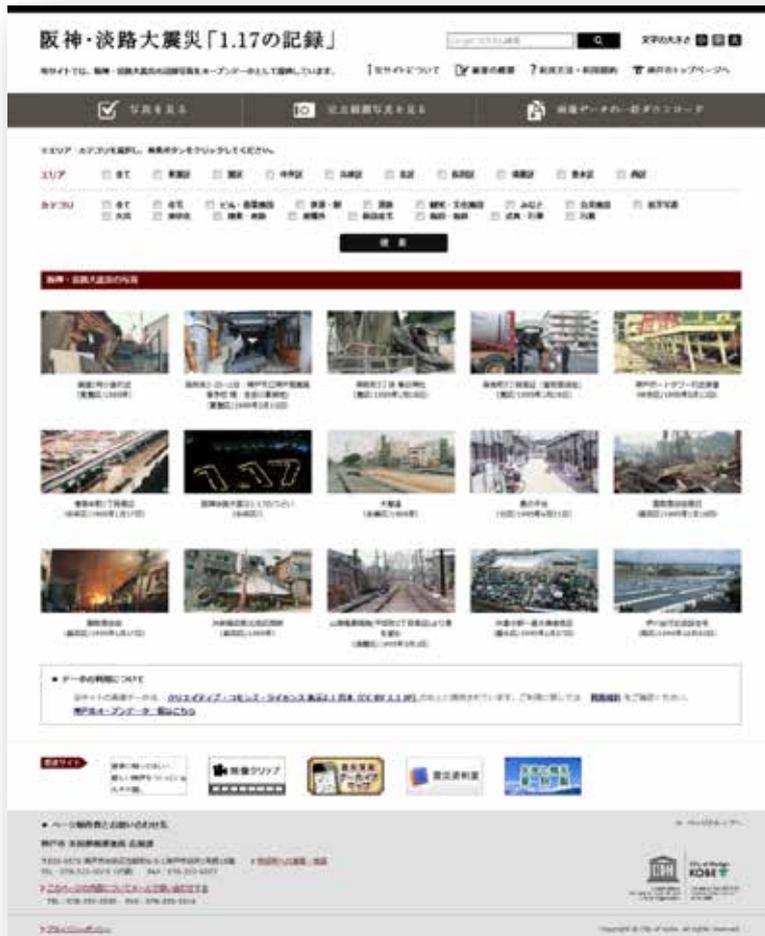
- 古地図を使って司法省を探したが、解釈が色々出てきた。
- 左側の岬を合わせてみると、右側の地形が合わない点に気づく。
- 最終的には、全く違う方向で、ピッタリと合う地点が見つかった。

景観を「読む」力



- 景観において、**不変な特徴に基づき位置合わせし、可変な部分の変化を見る。**
- 景観の不変な特徴を発見するには、**地形や文化的景観に関する背景知識が必要。**

災害アーカイブ



- **写真のオープンデータ**を用いて、災害復興の把握を助ける。
- 「**現場に行く**」ことが、災害の記憶の強化に有効なのは？
- **災害の集合的記憶**を再生、伝承するツールとして活用。

<http://kobe117shinsai.jp/>

「写真」を再考する



1. **一般的な定義**：写真（真を写す）とは「世界の記録」
写真を実世界に重ねるAR（拡張現実）の世界観。
2. **我々の定義**：写真とは「**撮影者による世界の記録**」
写真をファインダーに重ねるメモグラの世界観。

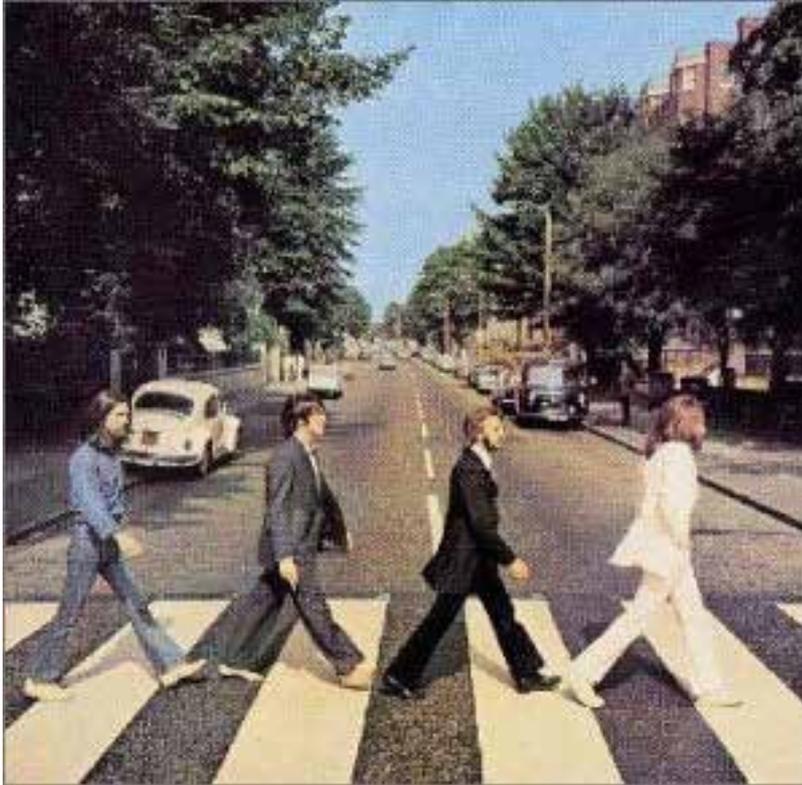
身体性と再体験



- 写真とは、撮影者のまなざしを記録したメディアである。
- 場所と姿勢を合わせ、時間を超えて撮影者の身体を再体験する。
- あの人は、なぜ、どうやって、この写真を撮ったのか？

https://twitter.com/third_sector_/status/490429114520567808

聖地化とコンテンツ・ツーリズム



- **ポップカルチャー**（映画・アニメ等）のシーンと同じ構図で写真を撮影する。
- 作品やスターと**身体性のレベル**で**一体感**が得られる。
- 多人数が訪れることで場所が「聖地化」。

アビー・ロード、ザ・ビートルズ
（画像はamazon.co.jpより）

断片化と再構成

遷画：展示の数学モデル に基づく参加型アーカイ ブの分析と ミュージアムでの展開

研究協力者：

神田 涼（当時：筑波大学）

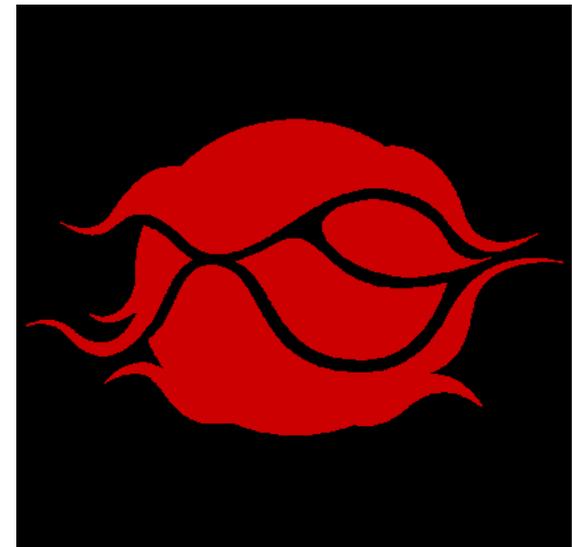
大西 磨希子（当時：国立情報学研究所）

西村 陽子（当時：国立情報学研究所）

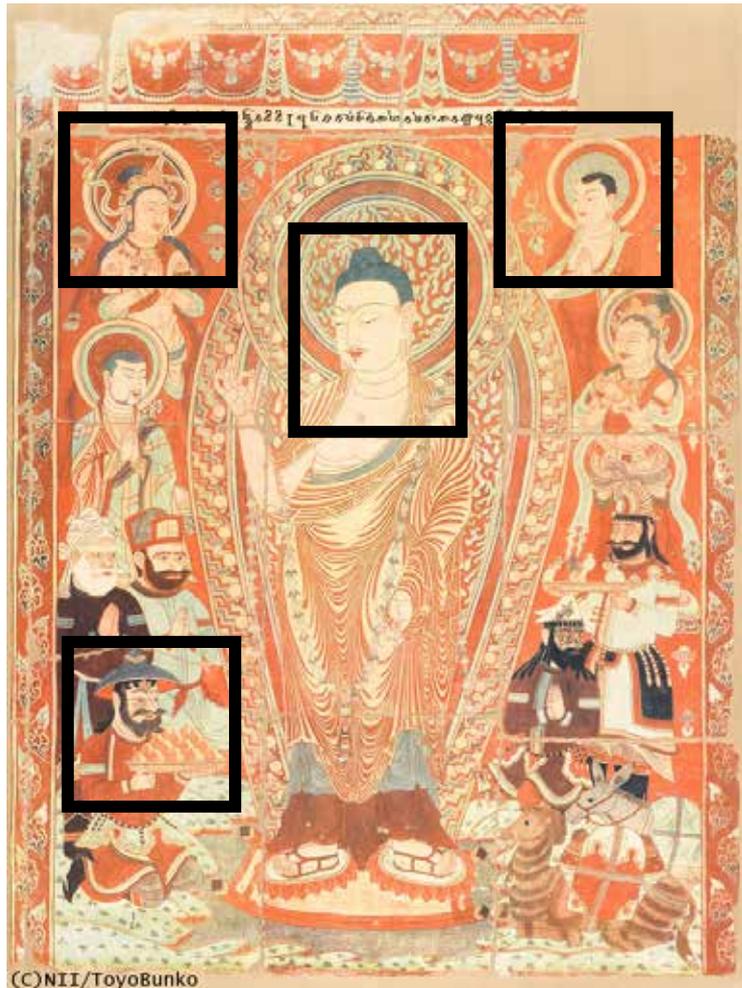
村松 賢子（当時：国立情報学研究所）

丸 奈美（国立情報学研究所）

公益財団法人 東洋文庫

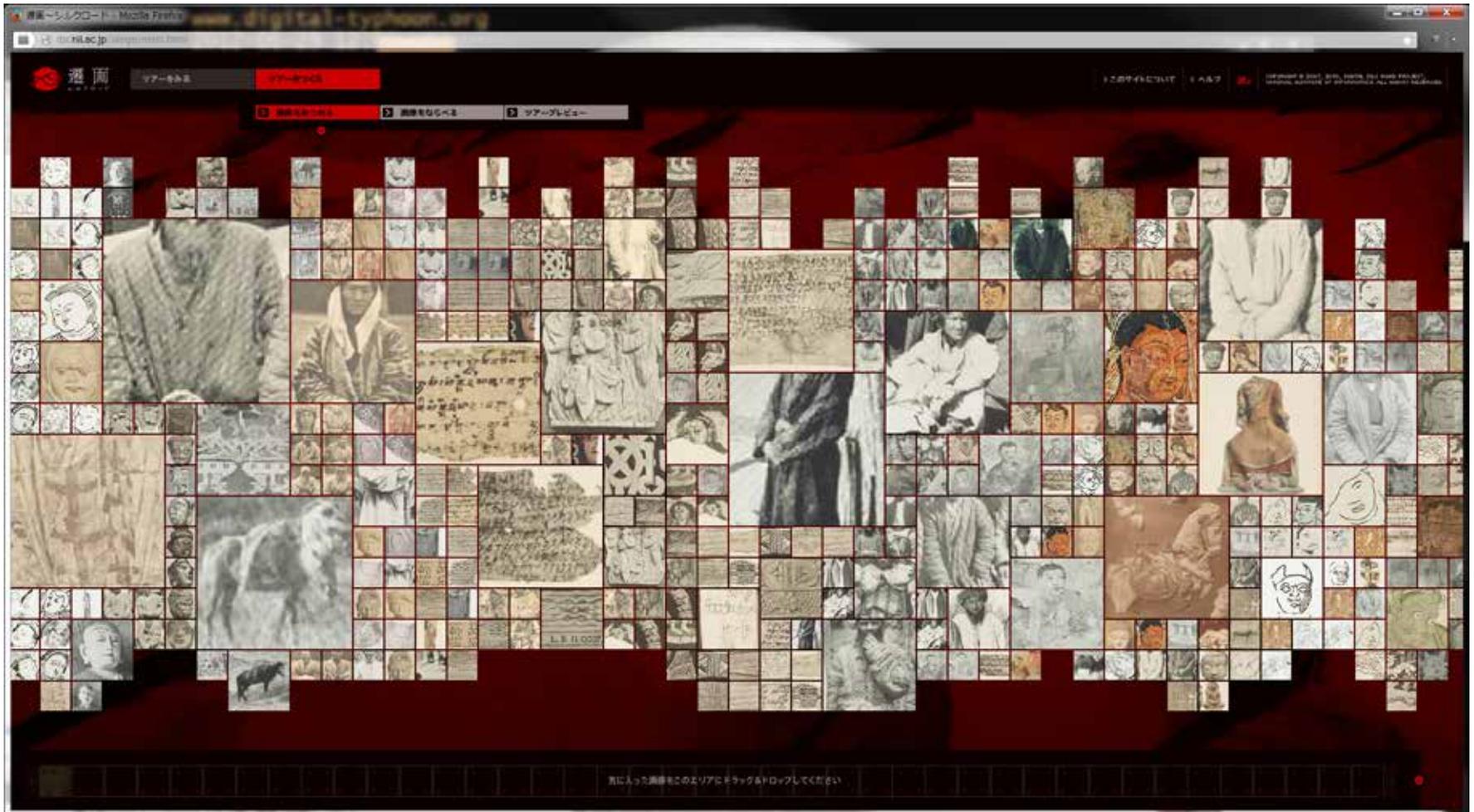


作品の断片化と素材化



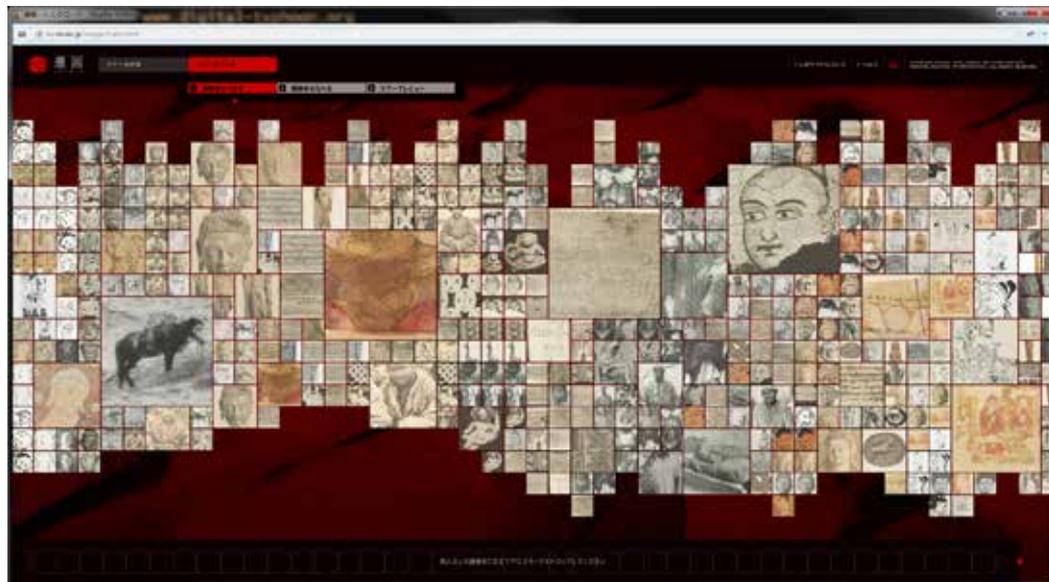
- 東洋文庫貴重書デジタルアーカイブでデジタル化した書籍。
- デジタル化した画像の一部を切り出し、素材画像として利用。
- 多数の作品を断片化し、メタデータを付与してコレクション。

素材の再構成とキュレーション



「遷画」の基本構造

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



収集支援機能

テーマ検索
エリア検索
類似画像検索
協調画像検索

収集 (Collect)



並べ替え (Order)



共有



ツアー展示
ポストカード



COPYRIGHT © 2007, 2010, DIGITAL SILK ROAD PROJECT,
NATIONAL INSTITUTE OF INFORMATICS. ALL RIGHTS RESERVED.

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

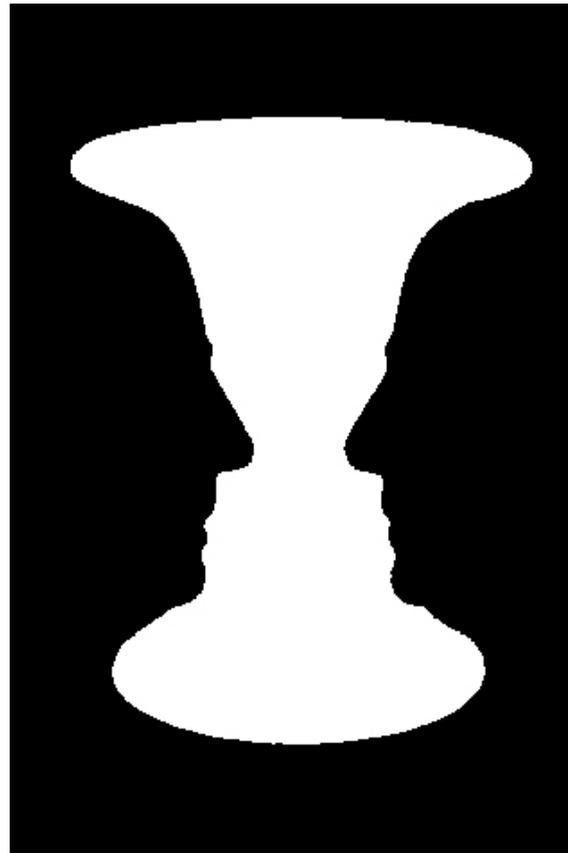
東洋文庫ミュージアム来館記念
2012年03月08日
ONICO
イケ面

ワークショップの開催



- 子供たちが遷画を体験しながら作品制作。
- 作品のテーマは自らが考えて設定する。

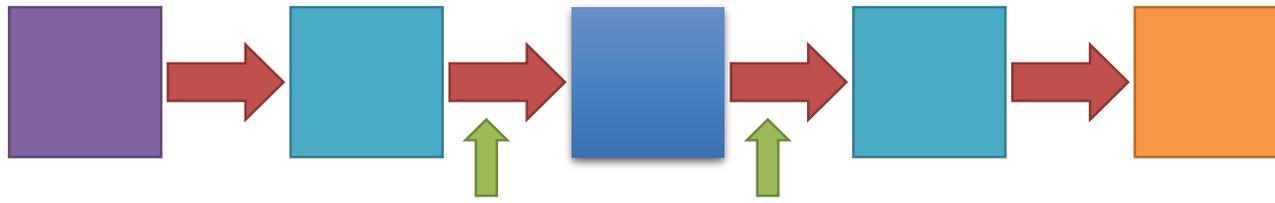
画像の多義性



ルビンの壺

出典：Wikipedia

断片化と再構成



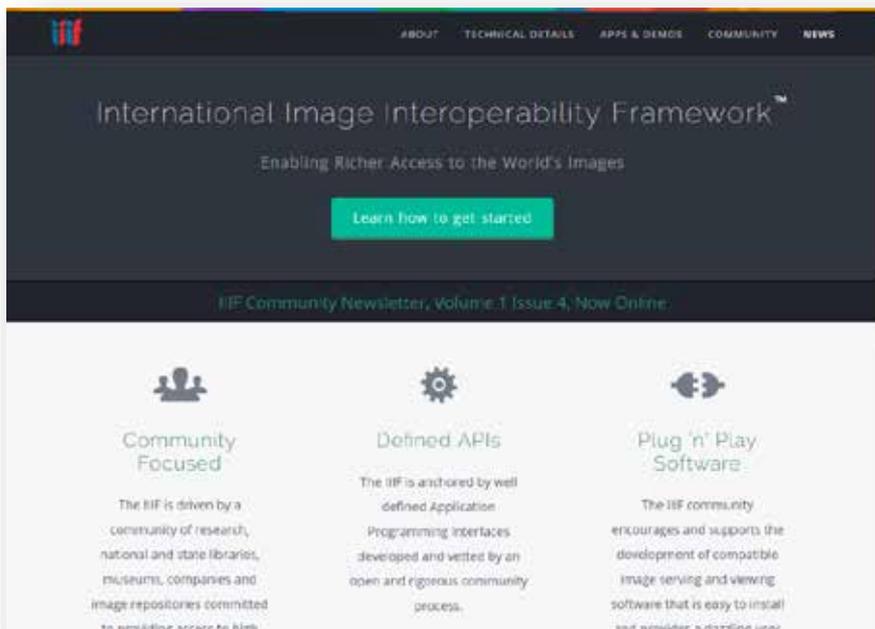
ローカル = 集合的な文脈が多様な遷移を生成する



グローバル = 局所的な構造だけでは統一感が生まれない

- いったん断片化した文脈を、マルコフ連鎖で再構成すると、新しいツアーが自動生成できる。
- 優れたキュレーションには全体像がある。グローバルな意味生成のモデルが必要。

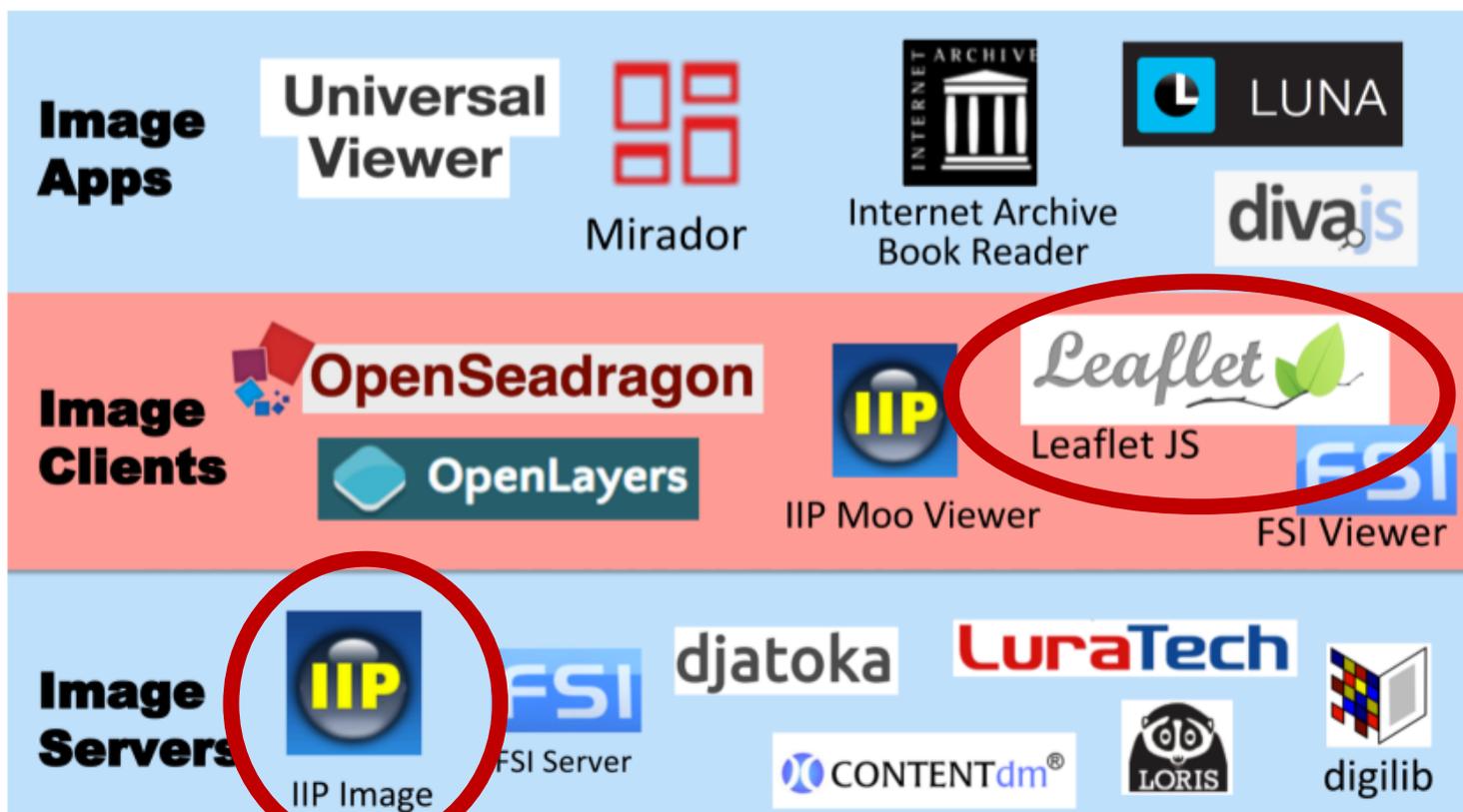
IIIF (International Image Interoperability Framework)



<http://iiif.io/>

1. 世界中のミュージアムやライブラリ、大学等に広まりつつある、**画像公開（配信）の標準技術**。
2. **すでに3億5千万件以上の画像がIIIFでアクセス可能**。
3. **よい検索方法がない**。Yahoo!以前の状況。

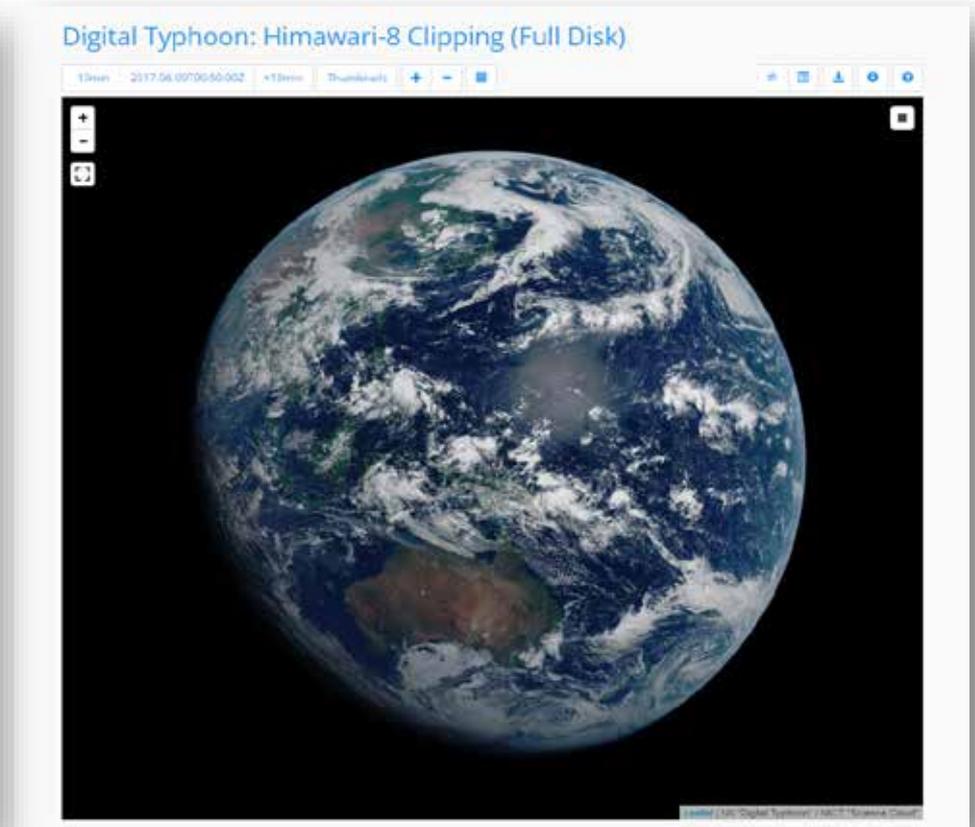
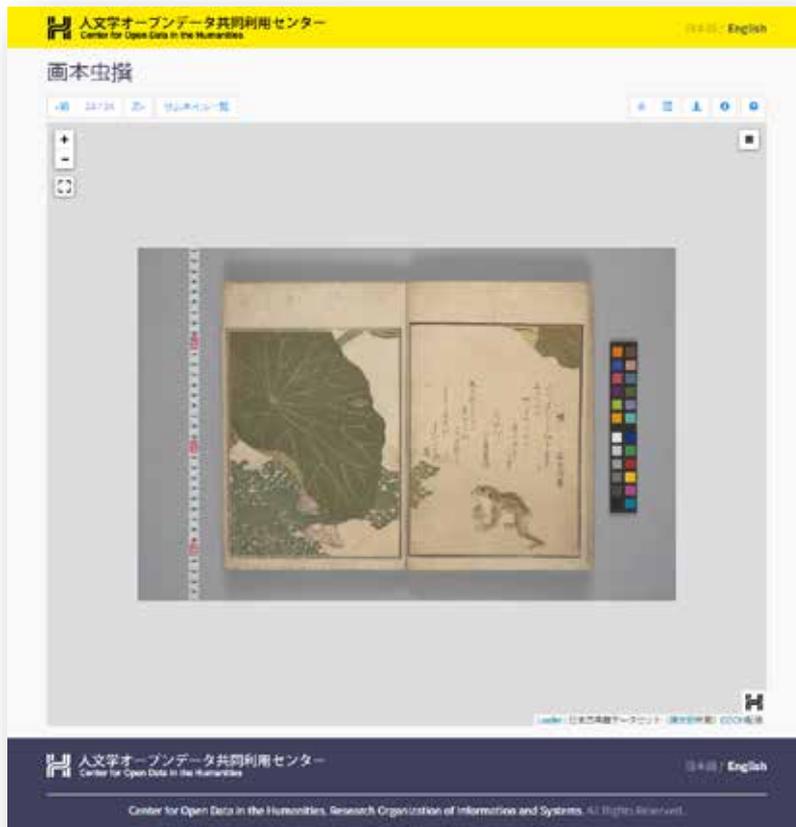
Commercial & Open Source Communities



Sheila Rabun, IIF Community Groups & Engagement, IIF Conference 2017.

IIIF Curation Viewer (for Timeline)

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

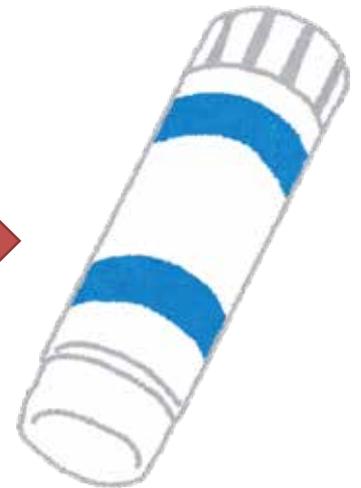
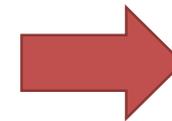
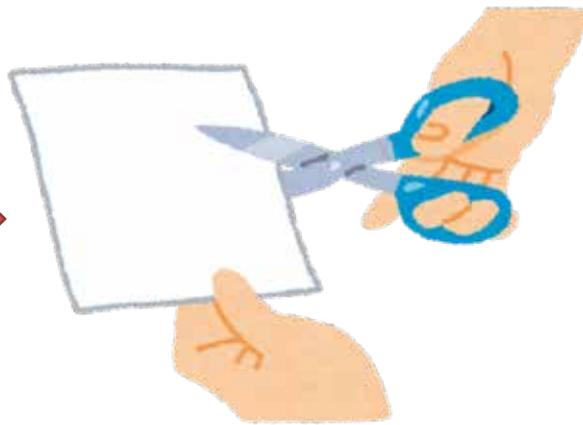
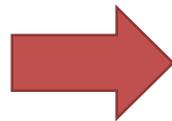


古き良きアナログ世界



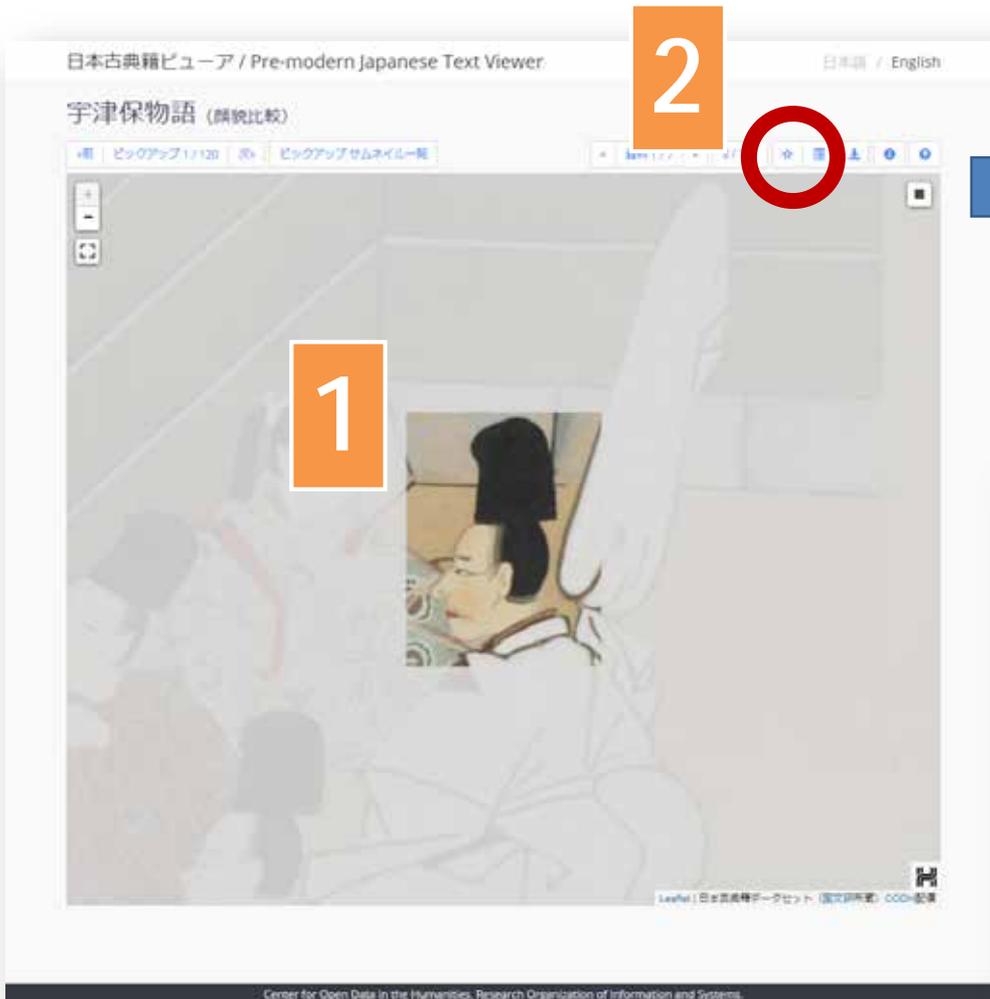
1 ハサミ

2 ノリ



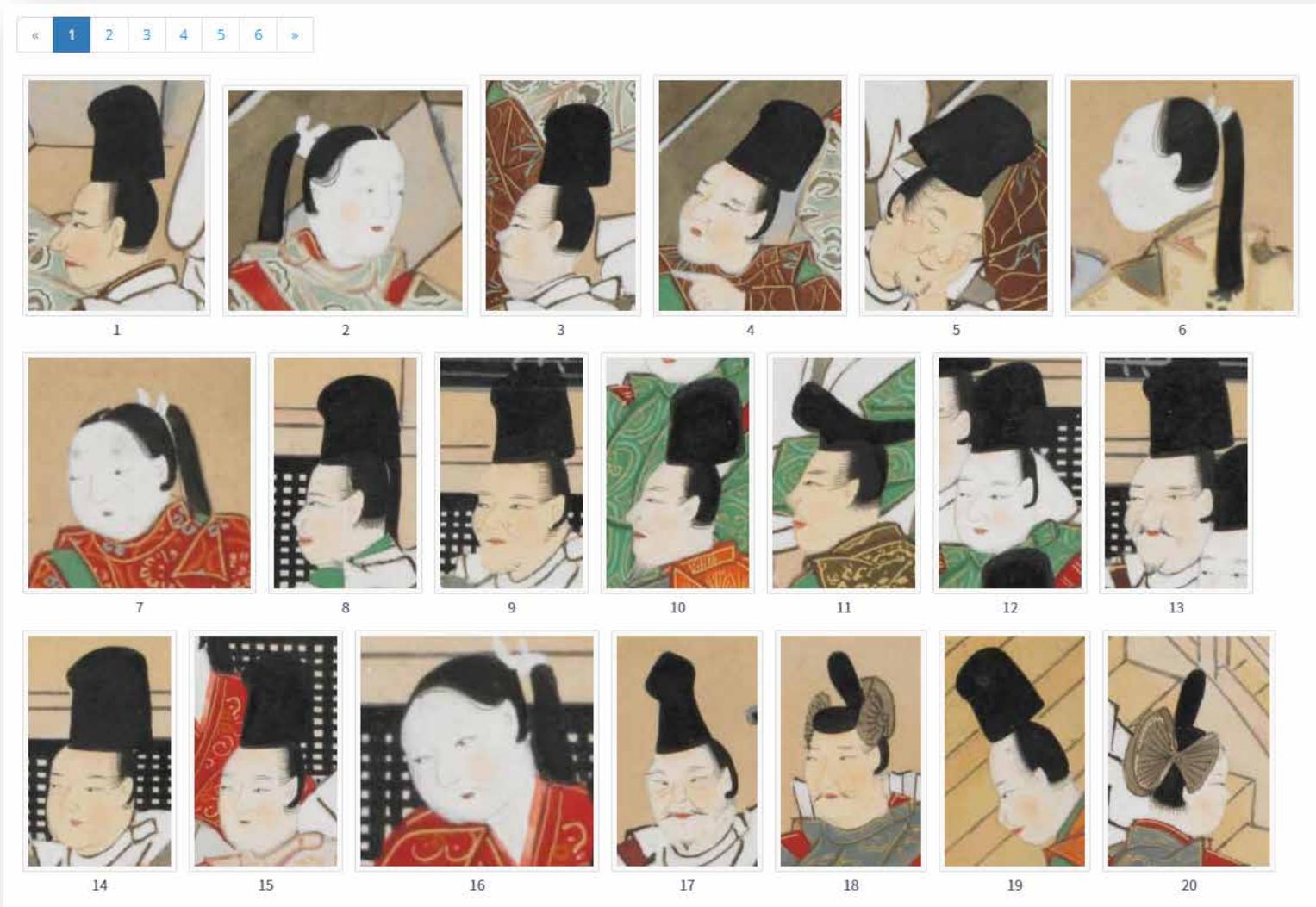
出典：いらすとや、<http://www.irasutoya.com/>

摩擦なきデジタル世界



「1. 四角で囲んで」「2. お気に入りに加える」だけ





顔貌比較 <http://codh.rois.ac.jp/pmjt/curation/3/>

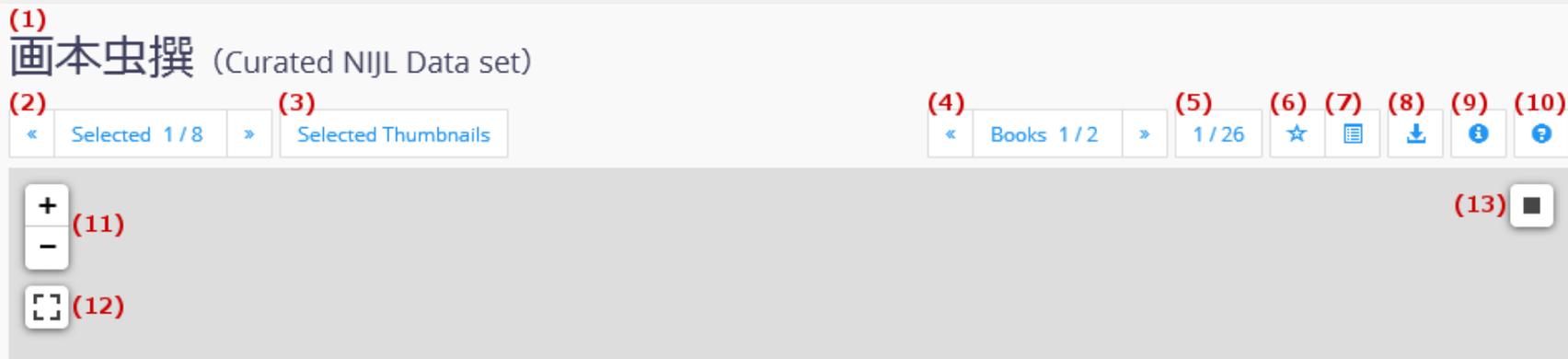
『宇津保物語』日本古典籍データセット（国文研所蔵）CODH配信

IIIF仕様の拡張

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

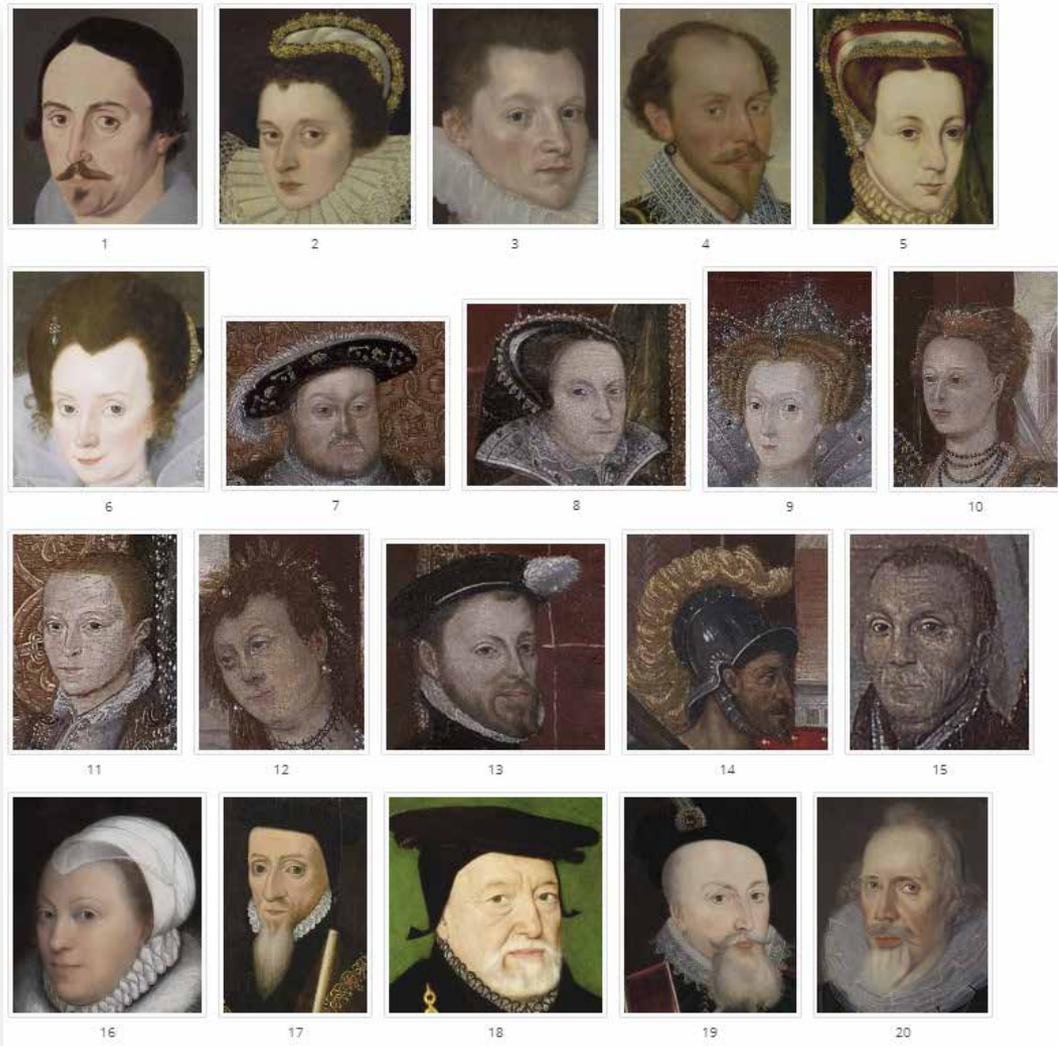
1. **Curation API** : テーマごとに資料横断的に画像を切り取り、整理し、展示する、という研究のニーズに応える。
2. **Timeline API** : IIIFが前提とする書籍のモデルを、科学画像で重要な時系列画像などのユースケースに拡張する。
3. **Cursor API** : 長大な系列に対し、CanvasのリストをManifestから一括取得せず、リストの一部を任意にリクエスト可能とする。

画像収集とキュレーション

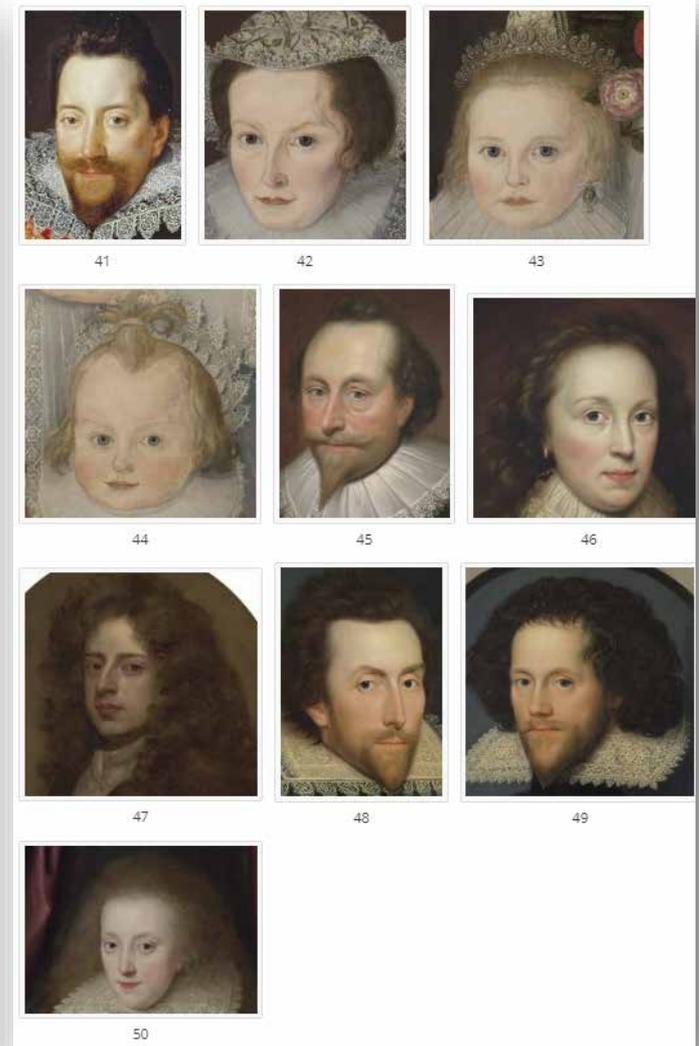


1. ' (6) は「お気に入り」ボタン。好きな画像を収集できる。
2. ' (13) は「切り取り」ボタン。四角で囲んで部分を切り取れる。
3. 2つの機能を組み合わせれば、**世界中の画像を集めて共有する仕組みが構築できる。**

16世紀



17世紀



顔貌データ：イギリスの肖像 <http://codh.rois.ac.jp/curation/exhibition/2/>

Originally from Yale Center For British Art

オープンサイエンス

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

オープンデータとデータ共有



**オープン
データ**

誰でも（制限なく）アクセスできる。

データ共有

コミュニティ内ではアクセスできる。

**クローズド
データ**

ごく限られた人のみアクセスできる。

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」

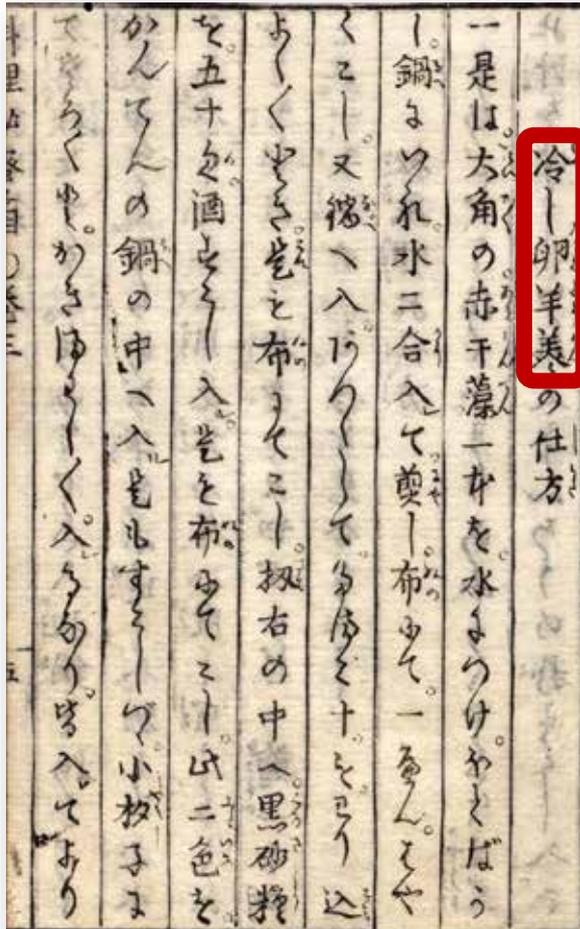


古いものに新しいものを継ぎはぎして、うまくやろうとしていないか。これはデジタルアーカイブにも通じる問いかけではないか。

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>

市民のためのオープンデータ

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

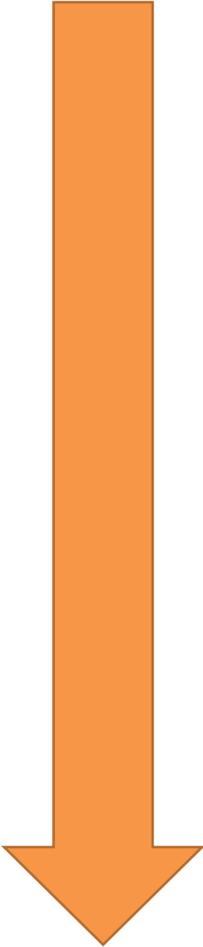


日本古典籍データセット
(国文研所蔵)



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

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。

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

●開催要旨 ●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(受付は午後)

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

主催 人文文化研究機構 国文学研究資料館

共催 学術文化研究機構 国文学研究資料館

後援 学術文化研究機構 人文科学センター 国文学研究資料館
日本学振興会・ヒューマン・エッセイ・センター

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

関連サイト 国立情報学研究所「情報学研究所データリポジトリ」
①「江戸時代食文化データベース」
②「江戸時代食文化データベース」
<http://www.nijl.ac.jp/cijproject/>

【開催要旨】
国文学研究資料館では、歴史的なアイデアをもとにする方法には、画像データや音声データの活用を
考え、進めています。国文学研究資料館の歴史的なデータと音声データ約300点(全冊約3
万2千冊)を国立情報学研究所(NII)のご協力のもと、国文学研究所「情報学研究所データリ
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。2015年11月16日からデータセットとして公開。アイデアソンではこのオープン
データを活用し、
国文学研究所「江戸時代食文化データベース」に公開されたアイデアを出し合い、おもてなしの
アイデアの発表、参加型のイベント。

【プログラム】
①15:00～15:30 歴史的典籍オープンデータとは
15:30～16:00 国文学研究資料館に関する大型プロジェクトの概要
16:00～16:30 オープンデータ入門(1) (45分/分組)
16:30～16:45 質疑応答・アイスブレイク
(15分/15分/分組)
②16:45～17:00 休憩
③17:00～17:15 歴史的典籍アイデアソン
17:15～17:30 オープンデータ入門
17:30～17:45 休憩
17:45～18:00 アイデアソン・セッション2
18:00～18:30 休憩
18:30～19:00 発表・まとめ
※プログラムは当日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)に限定して申し込みをさせていただきます。

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。国文学研究所の画像データと音声データ約350点(全冊約6万4千冊)を国立情報学研究所(NII)のご協力のもと、国文学研究所「情報学研究所データリポジトリ」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。



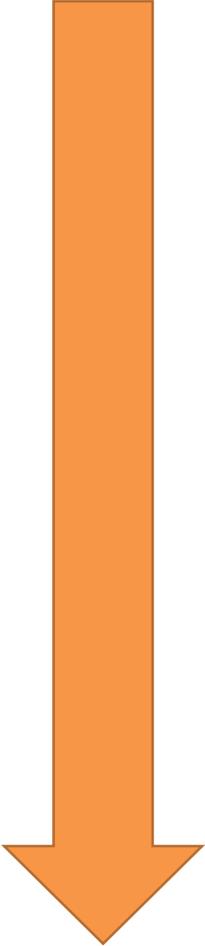
もちろん元となった古典の資料やその翻刻文、さらには現代語訳や挿絵などもウェブにあげ、一つの料理から多彩な情報や知識が広がる仕掛けを凝らし、SNS経由で「いいね！」を押してもらおうなどの工夫も必要。また、料理は地域ごとに違いがあるので(たとえばお雑煮の多彩さなど)、そうした情報をウェブにあげて比較できるようにすることも面白いと考えた。更に、古典を紹介した上で、日本料理の情報を発信することで、

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

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

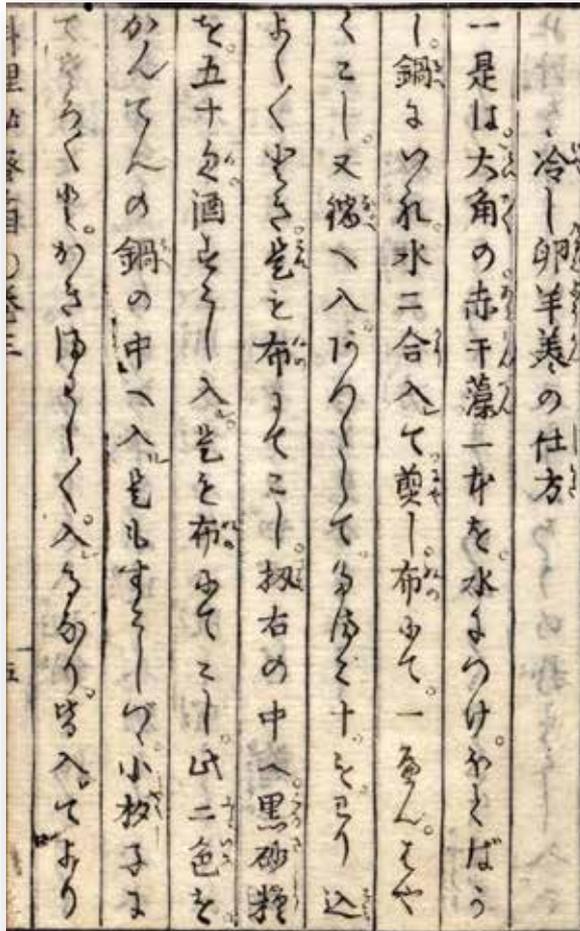
1. 江戸の料理本をデジタル化



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

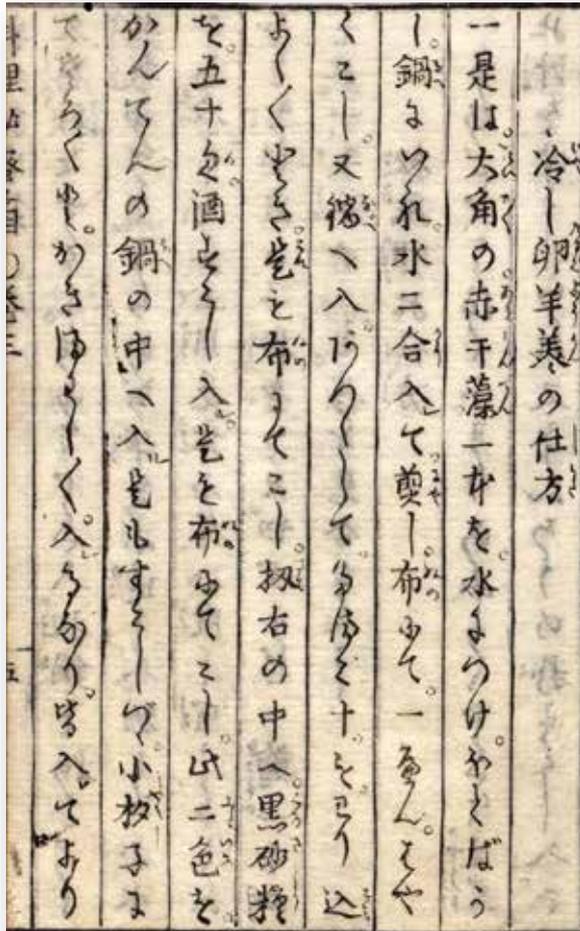


日本古典籍データセット
(国文研所蔵)

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

江戸料理レシピデータセット (CODH制作)

3. 翻刻を現代語訳

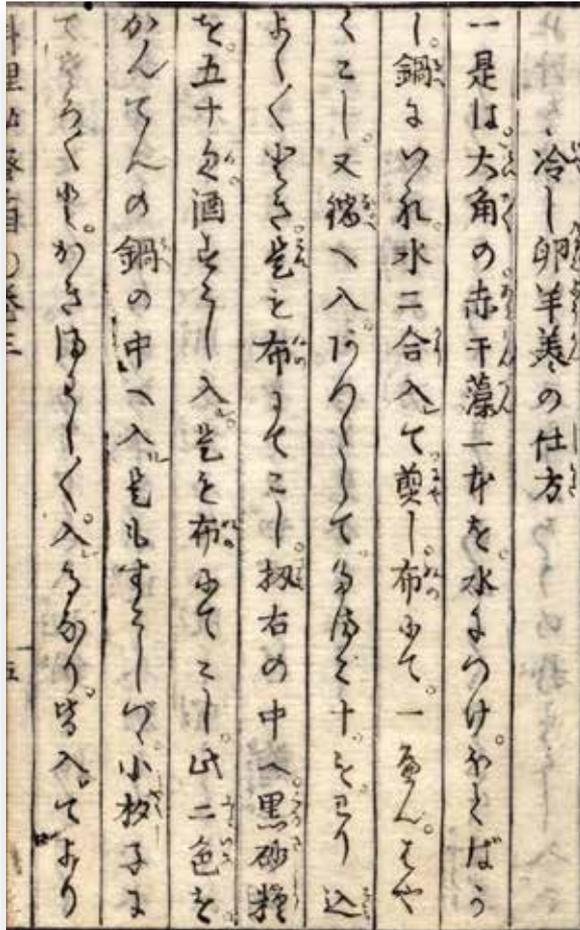


日本古典籍データセット
(国文研所蔵)

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



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。

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の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で公開。

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

人々の典型的な反応



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）

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

検索: 料理名・食材名

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. 寒天を水につけて、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. 3に入れ、再びこします。

5. 鍋に寒天と水(180cc)を入れて煮とがします。

6. 6を布などでこし、再び鍋に入れて熱します。

7. 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>

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

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

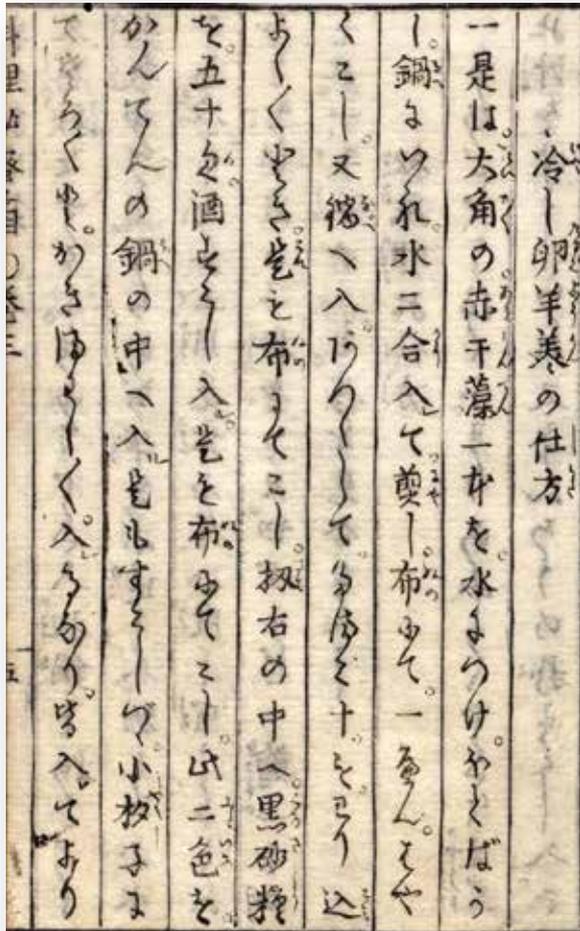


1 1,074 971

1052 リツイート

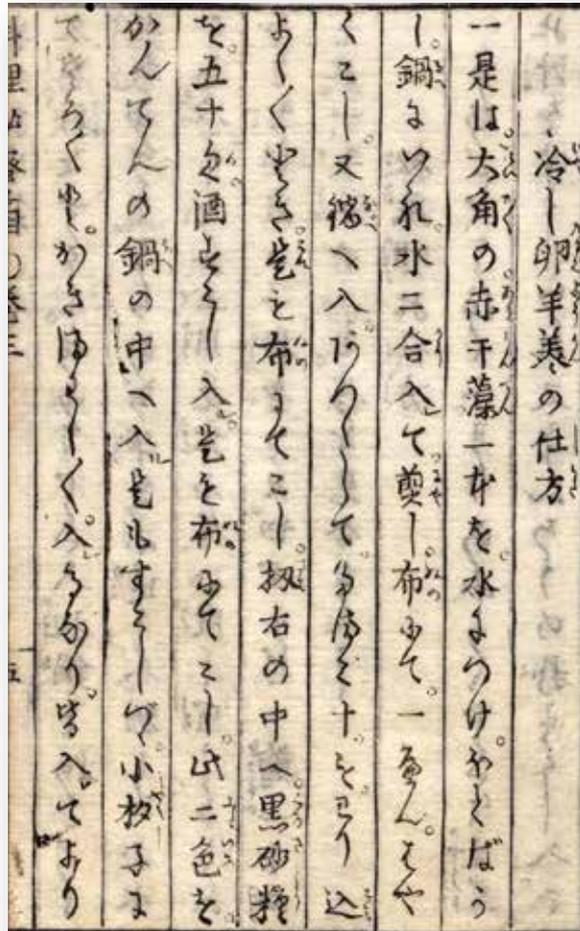
<https://twitter.com/jouhouken/status/801693251052781568>

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？

cookpad 253万レシピ

江戸時代のスイーツ 甘さスッキリ冷羊羹

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

卵	5個
寒天(赤)	1本(4g)
黒砂糖	100g
水	180cc
片栗粉	適量
酒	適量

カロリー: 271kcal/人 @ 塩分: 0.3g/人 @

1. 寒天を水につけて、ふやかします。

2. 生卵をよく混ぜます。

3. 冷めた生卵を布でこします。

4. 3に入れ、再びこします。

5. 鍋に寒天と水(180cc)を入れて煮ます。

6. 6を布などでこし、再び鍋に入れて熱します。

7. 7の熱した寒天の中に、5の卵液を少しずつ入れます。

クックパッドへの「展示」

The screenshot shows a recipe page on the Cookpad website. The title is '江戸時代のスイーツ 甘さスッキリ冷卵羊羹' (Edo-era Sweets: Sweet Sticky Cold Sheep Brain). The ingredients list includes: 餅 (5個), 寒天(赤) (1本(4g)), 黒砂糖 (100g), 水 (180cc), 片栗粉 (適量), and 酒 (適量). The page features a main image of the finished dish and a series of eight numbered steps with small images and text instructions for the preparation process.

- 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

人文情報学の可能性

- 人文学者は、デジタル技術の利活用によって、研究の問いを深められる。
- 情報学者が加わることで、抽象化と概念化が進み、横断的な研究が可能となる。
- デジタル人文学は情報技術（特にAI）の批判的研究（Criticism）も役割の一つ。
- AI時代の人機分業によって、量と質の両方を追究する研究を可能としたい。

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。

謝辞・参考資料

- 東洋文庫貴重書デジタル化では東洋文庫の協力を得ました。
- シルクロード遺跡研究は東洋大学の西村陽子氏との共同研究です。
- 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。
- より詳しい情報は、センターのウェブサイトをご覧ください。
- <http://codh.rois.ac.jp/>