

公開シンポジウム「総合知創出に向けた人文・社会科学
のデジタル研究基盤構築の現在」

画像公開方式IIFと歴史GISによる データ統合と総合知



北本 朝展（ROIS-DS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）

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

自己紹介

<https://researchmap.jp/kitamoto/>



@kitamotoasanobu

- 北本朝展（きたもと あさのぶ）
- 国立情報学研究所 教授、ROIS-DS人文学
オープンデータ共同利用センター セン
ター長
- 専門は、情報学、デジタル・ヒューマニ
ティーズ、データ駆動型サイエンス（地球
科学・防災等）など。オープンサイエンス
の展開に向けた、超学際的な共同研究の展
開にも取り組む。

ROIS-DS人文学 オープンデータ 共同利用セン ター (CODH)



深化

研究者

巨大化

機械



メンバー
国立情報学研究所
統計数理研究所
センター長＋
特任助教4名

市民

多様化

2016年4月～

「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



NIJL-NWプロジェクト

@国文学研究資料館

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

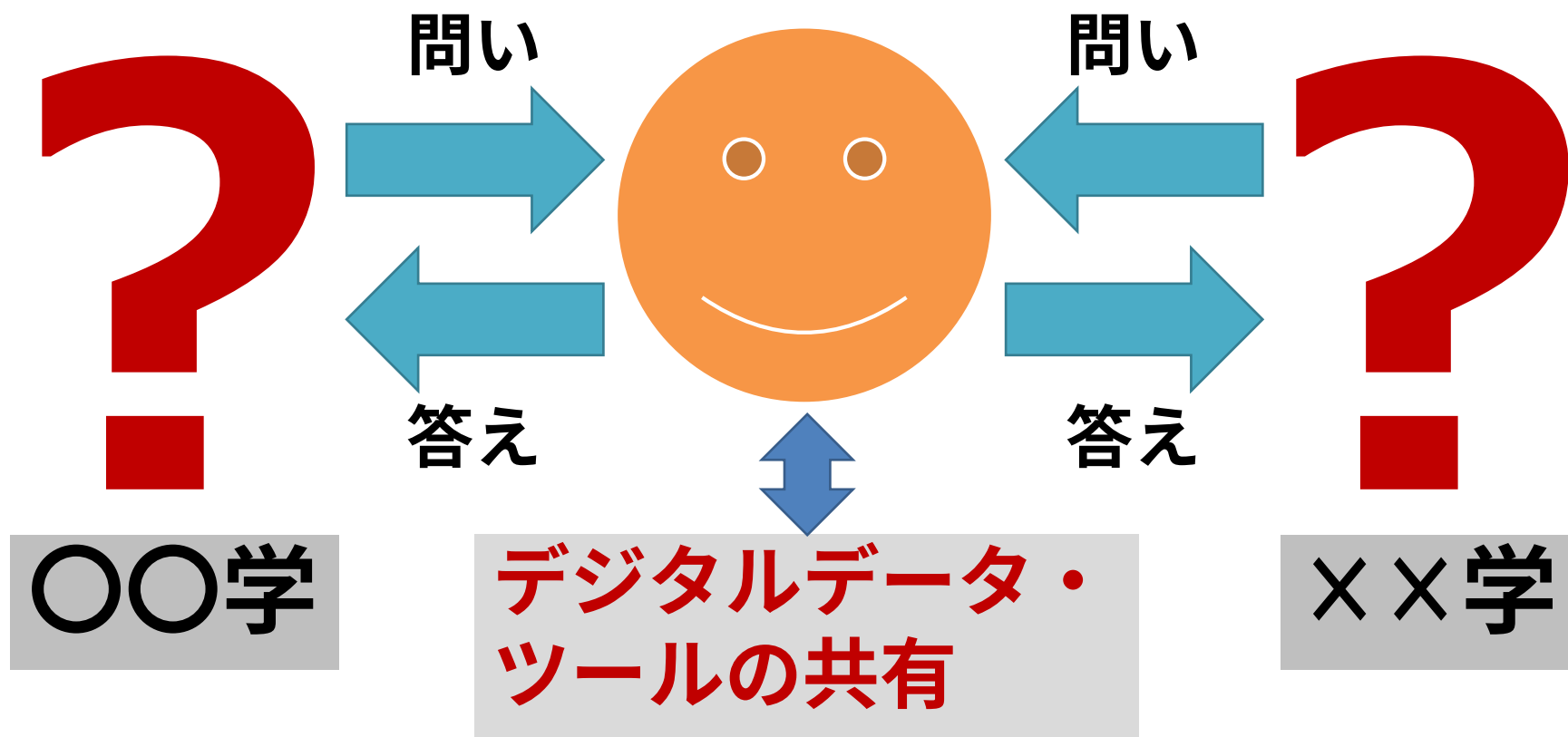
300,000点の日本古典籍
(1868年以前) を、デジ
タル化し、オープンデー
タとして公開。

日本文化のビッグ
データをどのよう
に活用するか？

デジタル人文学とは？

1. デジタル技術を人文学研究に導入することで、人文学の新しい研究方法を考案し、新しい知識を得ること。
2. デジタル技術の一つが「人工知能（AI）」。人間と機械の分業（協力）により、研究方法の変革が起こる。
3. コロナ禍が明らかにしたこと＝リモートでアクセスし、議論し、共有できることの価値も大きい。
4. 共同研究（チーム）が重要性を増す＝「ワンオペ」では研究資源が足りないため、協働の文化が鍵を握る。

多様な問いとデータの共有



IIIF（トリプルアイエフ）とは？

International Image
Interoperability
Framework = 国際的な
画像配信方式



Web : HTML
画像 : IIIF

IIIFサー
ビス1

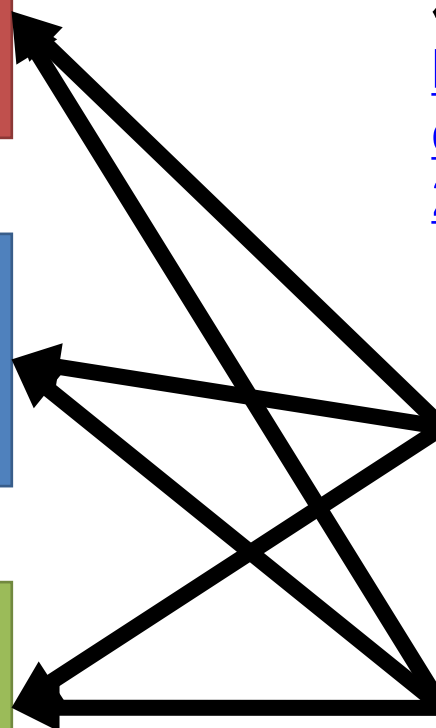
IIIFサー
ビス2

IIIFサー
ビス3

第14回CODHセミナー
IIIF Curation Platform
利活用レシピ100連発
<http://codh.rois.ac.jp/seminar/icp-recipe-20210218/>

IIIFビュー
ア1

IIIFビュー
ア2



IIIFを構成する技術要素

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

1. Image API

- 画像データにアクセスするための形式。

2. Presentation API

- 画像をまとめた本などの構造やメタデータにアクセスするための形式。

3. Authentication API、Content Search API、Change Discovery API、Content State APIなども仕様化。

4. 仕様に対応したオープンソース／商用ソフトウェアが多数存在するため、気軽に始められる。

Image API

<https://iiif.io/api/image/3.0/>

```
{scheme}://{server}/{prefix}/{identifier}/{region}/{size}/  
{rotation}/{quality}.{format}
```

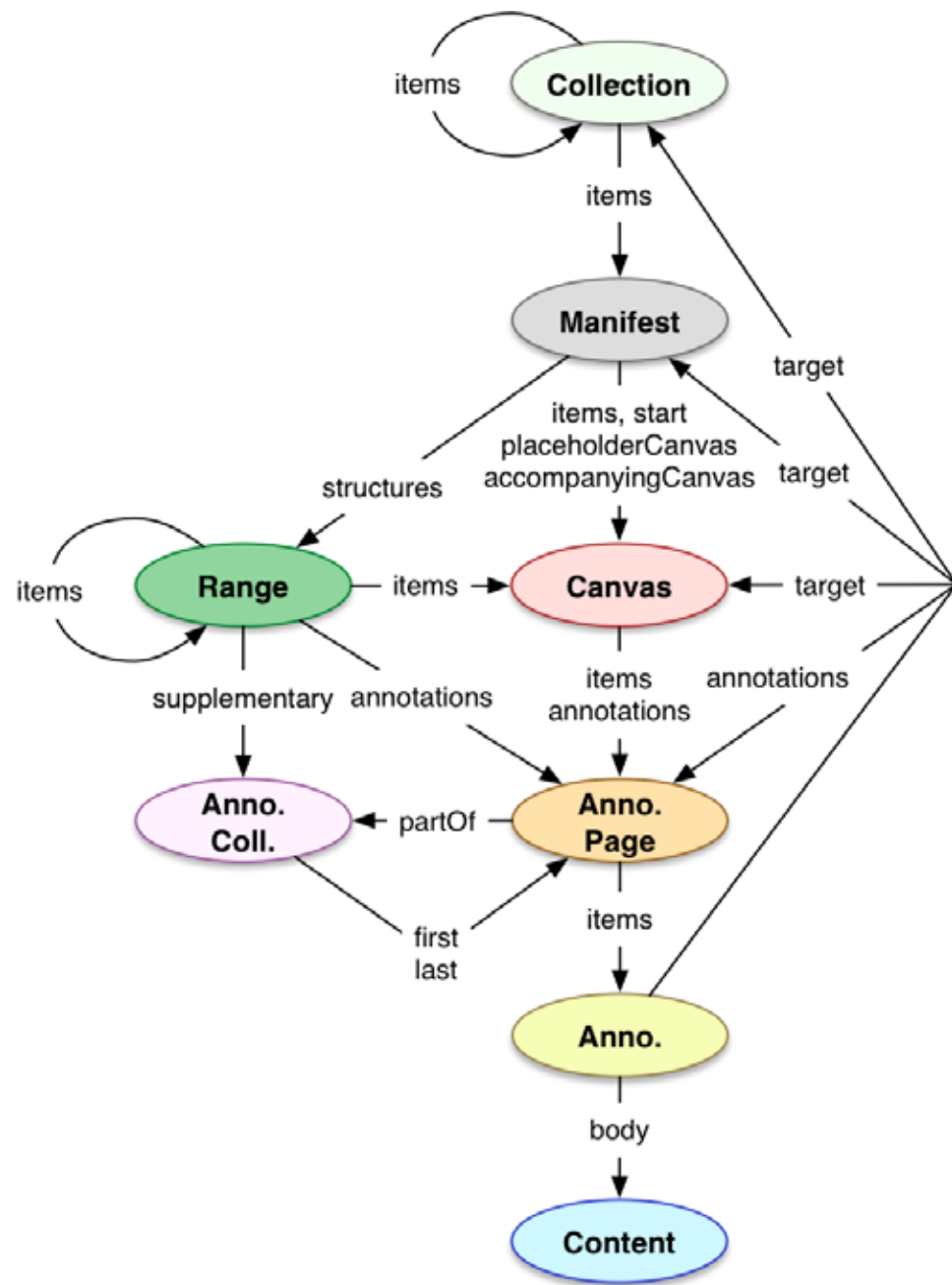
<https://example.org/image-service/abcd1234/full/max/0/default.jpg>

1. Example.orgという**ウェブサイト**の、image-serviceというprefixを持つ**サービス**の、abcd1234という**識別子**をもつコレクションの、full**領域**の、**最大サイズ**の、**回転**なしの、デフォルト**品質**の、JPEG**画像形式**のデータが欲しいというリクエスト。
2. 世界中のどのサイトも、同一の方法で画像にアクセスできる。

Presentation API

<https://iiif.io/api/presentation/3.0/>

1. マニフェスト (manifest) : トップレベルの構造。
2. キャンバス (canvas) : 一つ一つのページに対応。
3. アノテーション (annotation) : 様々な注釈を保持。
4. 全体はJSON形式で記述。
5. マニフェストをIIIFビューアに読み込ませて閲覧する。



IIIF Curation Platform (ICP)

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

キュレーションとは、もともと、ミュージアムにおいて、資料の収集や作品の展示などの活動を指すことばである。

1. あるテーマに沿って**コンテンツを集める**。
2. **適切な順番（配置）に並べる**。
3. **新たなコンテンツとして提示・共有する**。



1 ハサミ



2 ノリ



出典：いらすとや、<http://www.irasutoya.com/>

IIIF Curation Viewer

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

CODH開発



1. '■' の「切り取り」ボタン → 四角領域を指定
2. '☆' の「お気に入り」ボタン → 欲しい画像を収集



顔貌コレクション（顔コレ）

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

日本古典籍データセット（国文研蔵ほか）



IIIF Curation Viewer / IIIF Curation Platform
（CODH開発）



顔貌コレクション・データセット（CODH作成）



美術史研究



AI自動顔認識



機械の創造性
（GAN）

CODH 鈴木 親彦、東大 高岸 輝、Google Yingtao Tian、EPFL Alexis Mermetほか

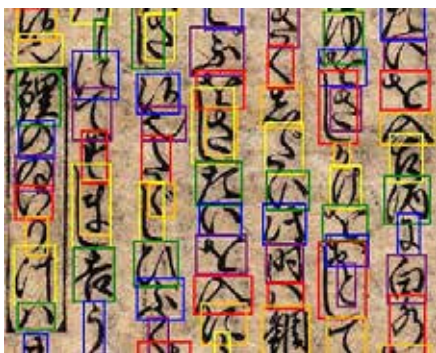
AIくずし字認識

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

日本古典籍
データセット
(国文研蔵)



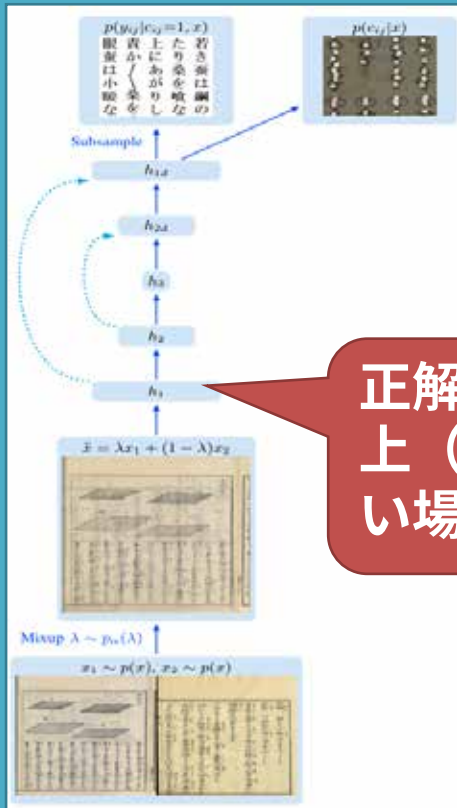
くずし字データ
セット (国文研・
CODH作成)



file	char	x	y
200003803_00024_2.jpg	U+3067	416	114
200003804_00024_2.jpg	U+3055	232	115
200003805_00024_2.jpg	U+304A	327	115
200003806_00024_2.jpg	U+3068	145	116
200003807_00024_2.jpg	U+3046	369	116
200003808_00024_2.jpg	U+305F	457	116
200003809_00024_2.jpg	U+5FA1	104	117
200003810_00024_2.jpg	U+3072	191	118
200003811_00024_2.jpg	U+540D	279	120
200003812_00024_2.jpg	U+3061	501	120

CODH カラーヌワット・タリンほか

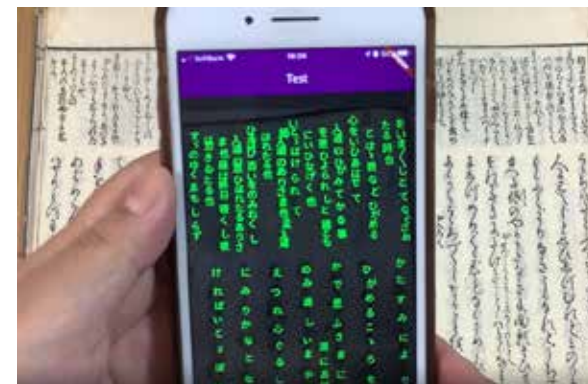
KuroNet
AIくずし字認識



くずし字認識サービス

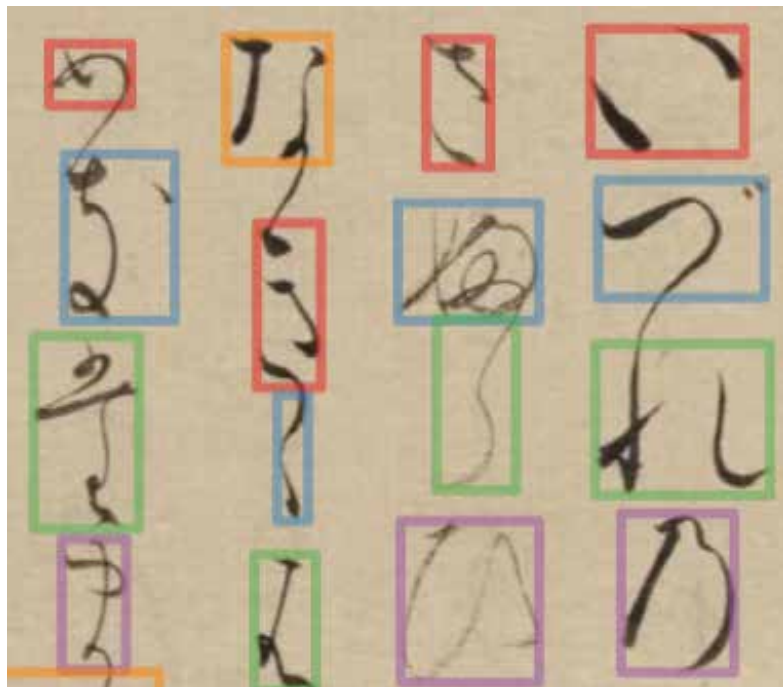
kaggle

くずし字認識コンペ



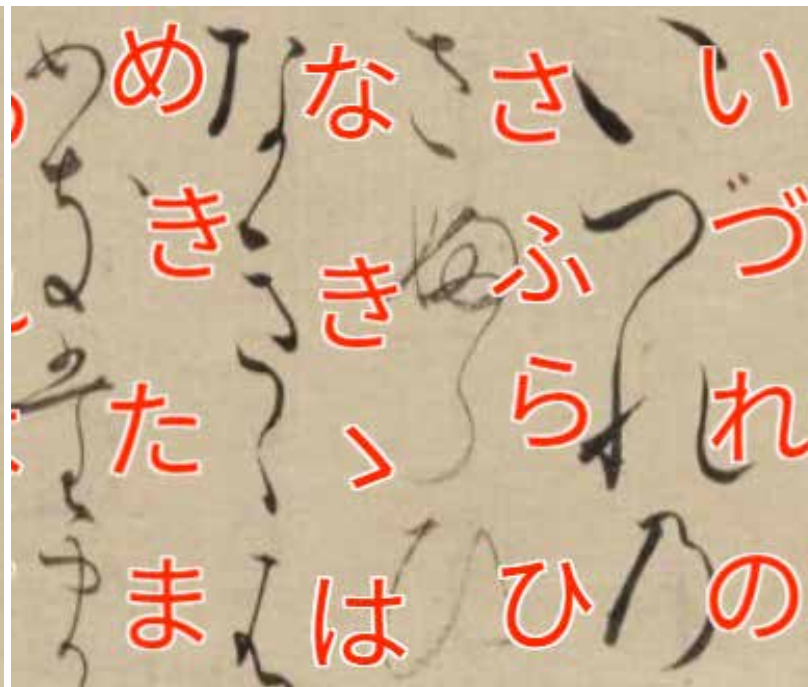
AIくずし字認識アプリ
「みを」

アノテーションビューモード（文字）

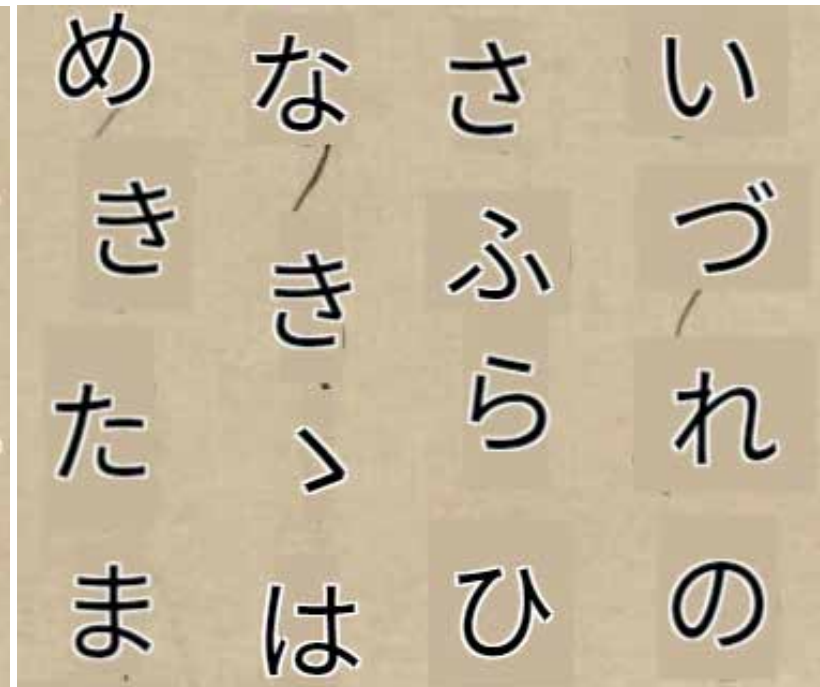


矩形枠線マーカー

『源氏物語』（国文研蔵）



文字マーカー（標準モード）



文字マーカー（上書きモード）

KuroNetくずし字認識サービス

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



- **くずし字多文字認識サービス。エクスポートされた画像に含まれるくずし字をすべて認識。**
- **1枚の認識は1秒～3秒程度。**

IIIFの相互運用性

1. IIIFの普及が世界中で急速に進んだ理由は、**画像提供者の共通のニーズにうまく答えたから**。
2. 共通形式でデータを公開する（相互運用性を担保する）ことで、**同じツールを使って世界中の画像にアクセスできるようになった**。
3. 各種ツールの前提がIIIF形式になると、**IIIF対応していないアーカイブは「不便」とみなされるようになる**。
4. **Fear of missing out（取り残される恐怖）が近づく**。

オープンな画像の利活用を開拓するIIIF Curation Platform, カレントアウェアネス-E, No.398, 2020.09.17,
<https://current.ndl.go.jp/e2301/>

歴史ビッグデータの統合解析

<http://codh.rois.ac.jp/historical-big-data/>

過去のビッグデータを統合解析するための基盤技術进行研究



歴史的資料
(史料)

自然科学的
データ

人文社会的
データ

気候

地震

噴火

疫病

経済

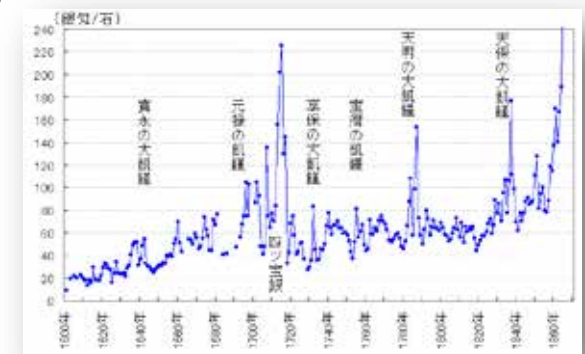
人口

政治

文化

データ
構造化
ワーク
フロー

歴史ビッグ
データ研究基
盤（機械可読
データ）



江戸マップ

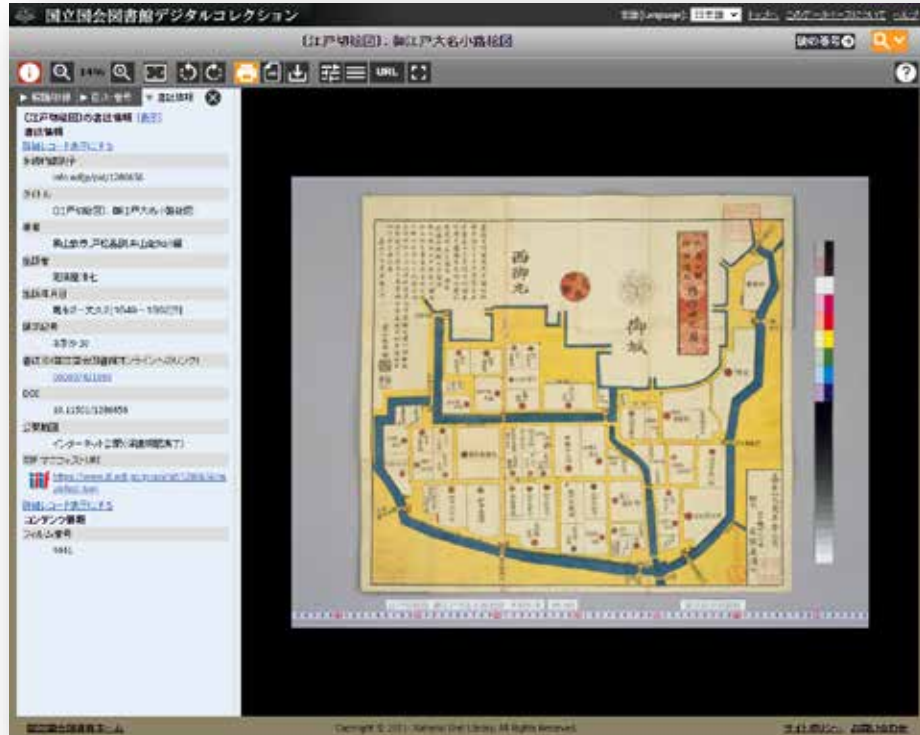
<http://codh.rois.ac.jp/edo-maps/>

番号	分類	現代語訳	翻刻	地図
2-001	施設	幸橋御門	幸橋御門	拡大図
2-002	施設	山下御門	山下御門	拡大図
2-003	施設	数寄屋橋御門	数寄屋橋御門	拡大図
2-004	施設	鍛冶橋御門	鍛冶橋御門	拡大図
2-005	施設	呉服橋御門	呉服橋御門	拡大図
2-006	地名	一石橋	一石橋	拡大図
2-007	地名	出橋	出橋	拡大図
2-008	町名	丸屋町	丸屋丁	拡大図

[2-296]
地名：磯辺大神宮（イソベ大神宮）
分類：寺社仏閣

29枚の地図から
8719ヶ所の地名を抽出して
データベース化

江戸切絵図の構造化



江戸切絵図で最も売れた尾張屋版（1849-1862）：国立国会図書館「デジタルコレクション」 doi:[10.11501/1286656](https://doi.org/10.11501/1286656)



IIIF形式だと、
画像をコピー
せずに、上に
注釈を重ねて
付加価値を付
与できる。



IIIF Curation Viewerで読み込み、地名
を矩形で囲んで画像座標を記録し、
矩形中の文字を翻刻・現代語訳し、
地名をデータベース化。

古地図のジオレファレンス



国立国会図書館
『江戸切絵図』

GCPの対応付け



立命館大学
日本版Map Warper



江戸マップβ版＋日本版Map
Warperタイル配信サービス



千代田区

© 2020 ZENRIN

Google Earth

2022/1/22

人文・社会科学のデジタル研究基盤構築

35° 40' 52.26" N 139° 45' 26.42" E 標高 2 m 高度 200 km

江戸買物案内 – 商業ビッグデータ

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



- 『江戸買物独案内』(1824) から、商人名や居所などを抽出しデータベース化。
- 歴博の「江戸商人・職人データベース」は現在の区レベル。
- 江戸買物案内は、江戸マップの町レベルでリンク。

江戸観光案内 – 観光ビッグデータ

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

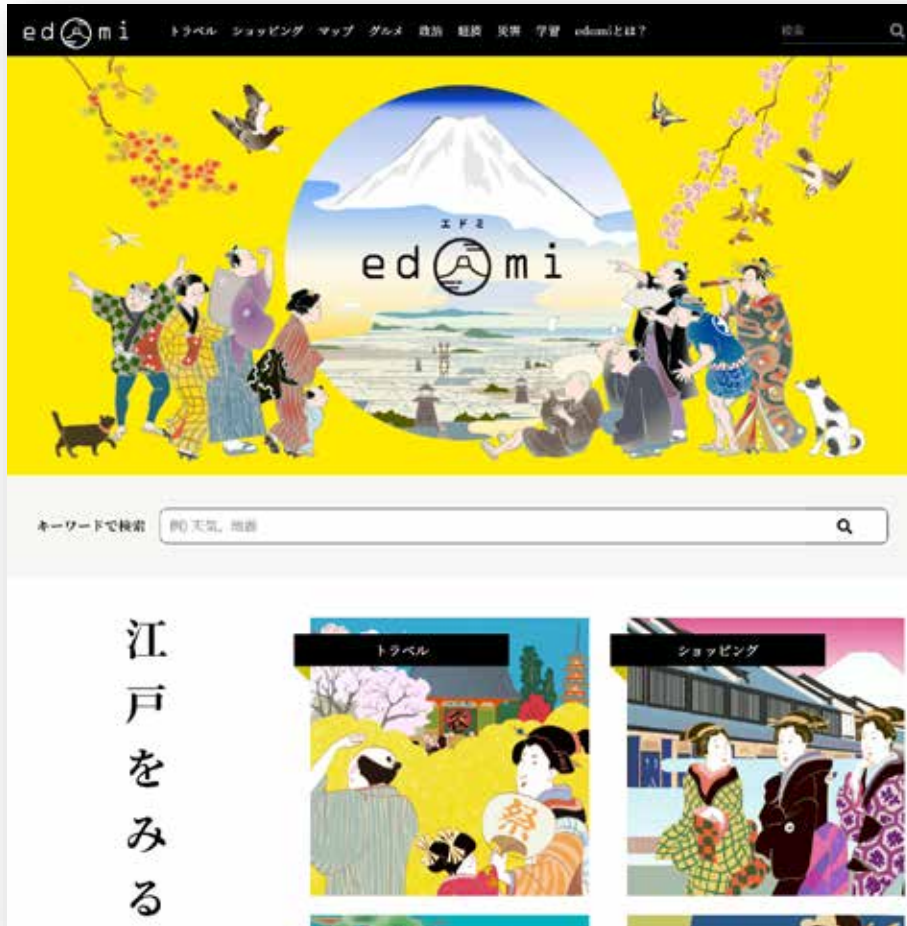


- 江戸時代の名所記や名所案内を各世紀2点ずつ選択。
- IIF Curation Viewerを用いて挿絵を切り抜き、地名を翻刻し、メタデータを付与。
- 江戸マップ、歴史地名データとリンク。

CODH 鈴木 親彦 ほか

edomi - 江戸をみる／みせるデータポータル

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



1. 江戸の都市、江戸の時代に関する**構造化データを公開するデータポータル**。
2. **現代の地図、現代の分類体系**など、現代の標準と過去のデータを接続する。
3. 現代のYahoo!のように、**カテゴリごとにデータを収集し、整理し、公開する**。

史料とエンティティのリンク



宇田川丁、三島丁・
神明丁、此分漬家多、
土蔵残所なし

固有表現認識

宇田川丁、三島丁・
神明丁、此分漬家多、
土蔵残所なし

御江戸大地震大破并出火類焼場等書上之写（みんなで翻刻）

1. 江戸マップに出現する地名を、**地名リソース（エンティティのデータベース）**として整備する。
2. **史料の文字列から固有表現（地名）を抽出する。**
3. **特定のエンティティとリンク**することで、実世界と紐づける。

曖昧性解消

原資料表記	江戸マップID	江戸マップ表記
宇田川丁	4-358	宇田川町
三島丁	4-290	三島丁
神明丁	4-294	神明町

歴史的行政区域データセットβ版

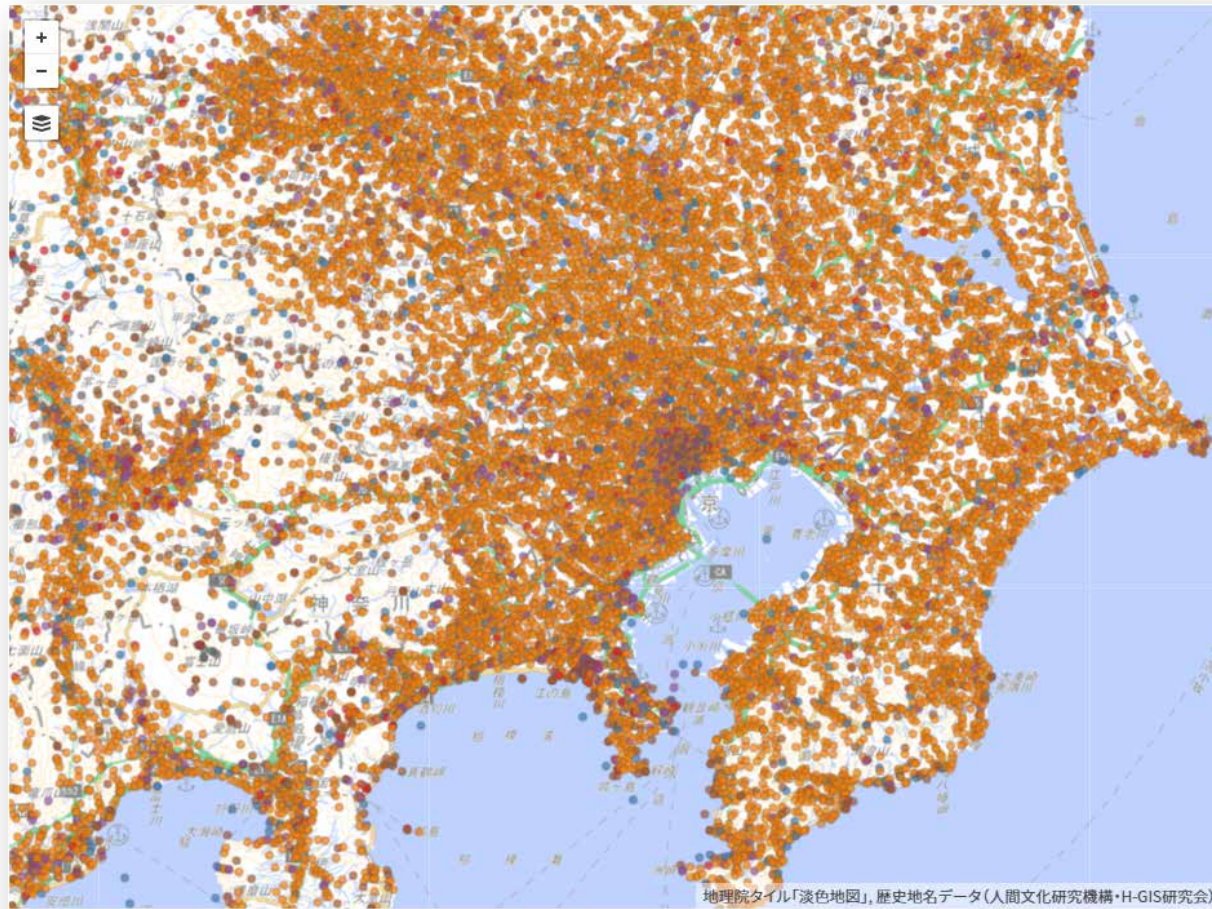
<https://geoshape.ex.nii.ac.jp/city/>



1. 1920年から2021年
まで、29時点。
2. 行政区域件数 =
16,447件、行政区域
境界データ件数
= 102,578件
3. 2017年1月に公開。
毎年データ更新と
機能追加を継続中。

歴史地名マップ

<http://codh.rois.ac.jp/historical-gis/nihu-map/>



マーカーの色は歴史地名データの属性に対応し、**行政地名**、**建物**、**水部**、**地形**、**名所旧跡**、**その他** とします。また

1. 人間文化研究機構・H-GIS研究会が公開する「**歴史地名データ**」298,914件を地図にまとめて表示。
2. **バイナリベクトルタイル**の活用により、多数地点をウェブ地図に表示。
3. **歴史地名データの識別子と属性**を、地名リソースの一つとして活用。

GeoLOD

<https://geolod.ex.ni.ac.jp/>

GeoLOD 地名情報を集約する地名情報処理システム

GeoLOD地名管理システム ログイン

川崎

検索

結果一覧 地名クラス 辞書

川崎 (市区町村) 岩手県/東磐井郡 岩手県東磐井郡川崎村

川崎 (市区町村) 宮城県/栗田郡 宮城県栗田郡川崎町

川崎 (市区町村) 宮城県/栗田郡 宮城県栗田郡川崎町

下川崎 (市区町村) 福島県/安達郡 福島県安達郡下川崎村

上川崎 (市区町村) 福島県/安達郡 福島県安達郡上川崎村

川崎 (市区町村) 福島県/西白河郡 福島県西白河郡川崎村

川崎 (市区町村/政令指定都市) 神奈川県/川崎市 神奈川県川崎市

川崎 (市区町村) 神奈川県/川崎市 神奈川県川崎市川崎区

川崎 (市区町村) 神奈川県 神奈川県川崎市

川崎 (市区町村) 神奈川県/横須賀市 神奈川県横須賀市川崎町

川崎 (市区町村) 岐阜県/本巣郡 岐阜県本巣郡川崎村

川崎 (市区町村) 静岡県/榛原郡 静岡県榛原郡川崎町

川崎 (市区町村) 三重県/鈴鹿郡 三重県鈴鹿郡川崎村

江川崎 (市区町村) 高知県/幡多郡 高知県幡多郡江川崎村

川崎 (市区町村) 福岡県/田川郡 福岡県田川郡川崎町

地名

GeoLOD ID: eN00cN

地名: 川崎

地名かな:

住所 (現在):

緯度: 35.887844

経度: 139.522042

地名クラス: 行政地名/字

上位語: 埼玉県

説明: 旧5万分の1地形図 大宮 (1910/08/30) 76-5-1

異表記:

出典: <http://codh.rois.ac.jp/historical-gis/nihumap/?id=40231674>

有効期限 (始点):

有効期限 (終点):

地名接頭辞:

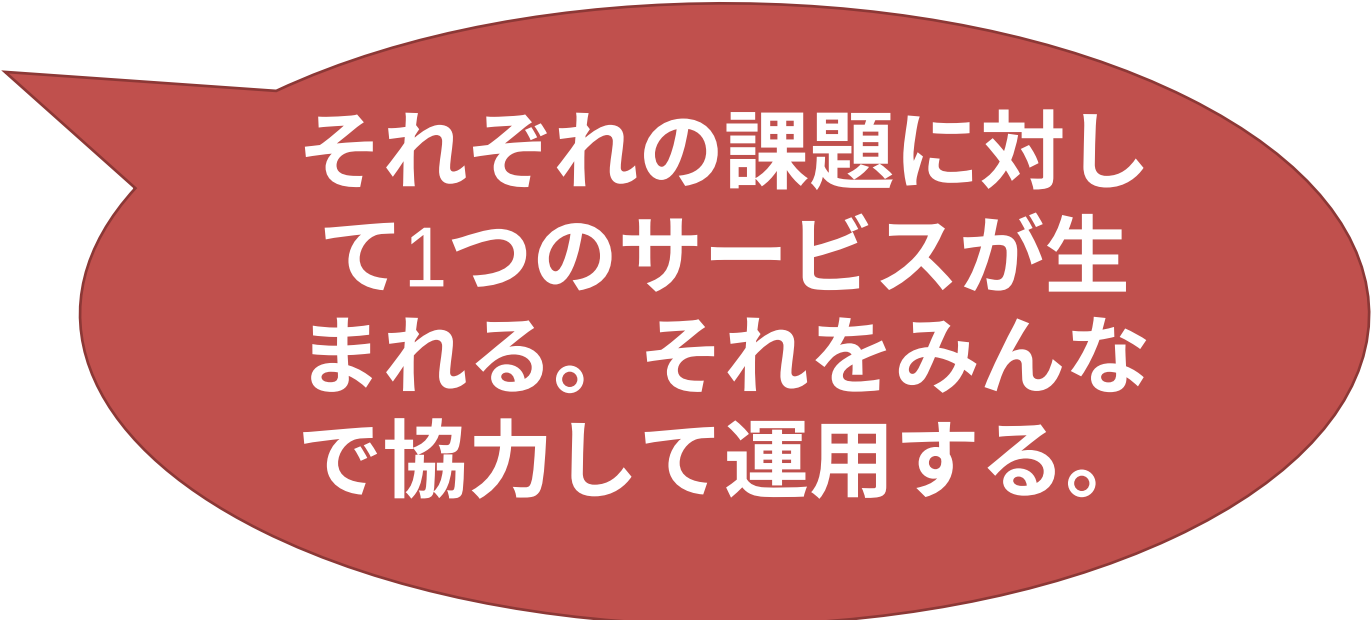
地名接尾辞:

周辺の説明はこちら

1. 登録した地名に**識別子 (GeoLOD ID)**を付与し、アプリを越えて共有するための**地名情報基盤**。
2. 作成した地名辞書は、**GeoNLPによる自動地名抽出**にも活用。
3. 有効期限など**歴史地名**にも対応。

1つのことをうまくやる

1. When → HuTime (国際日本文化研究センター 関野氏)
2. Where → GeoLOD
3. Who → ?
4. What → ?
5. Why → ?
6. How → ?

A red speech bubble with a white border, containing Japanese text. It points towards the list of questions.

それぞれの課題に対して1つのサービスが生まれる。それをみんなで協力して運用する。

UNIX哲学

<https://ja.wikipedia.org/wiki/UNIX%E5%93%B2%E5%AD%A6>

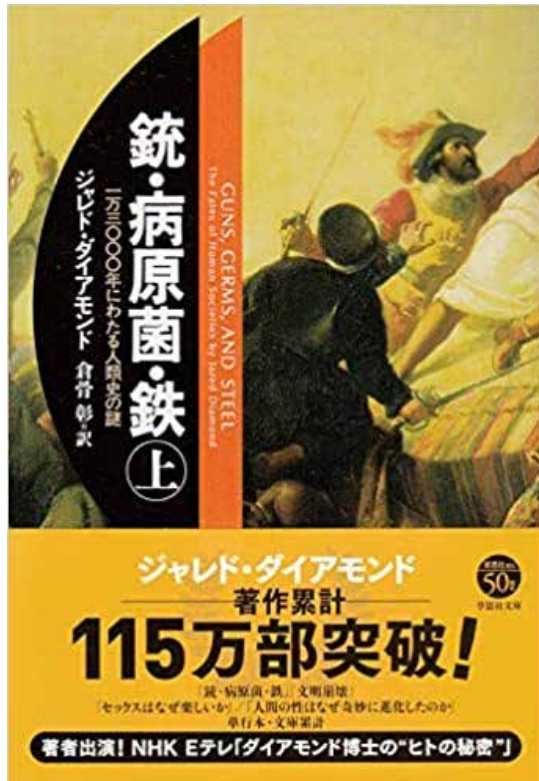
1. **一つのこと**を行い、
またそれをうまくや
るプログラムを書け。
2. **協調して動く**プログ
ラムを書け。
3. **Keep it Simple, Stupid**
(KISS原則、「シン
プルでつまらないも
のに保て」)



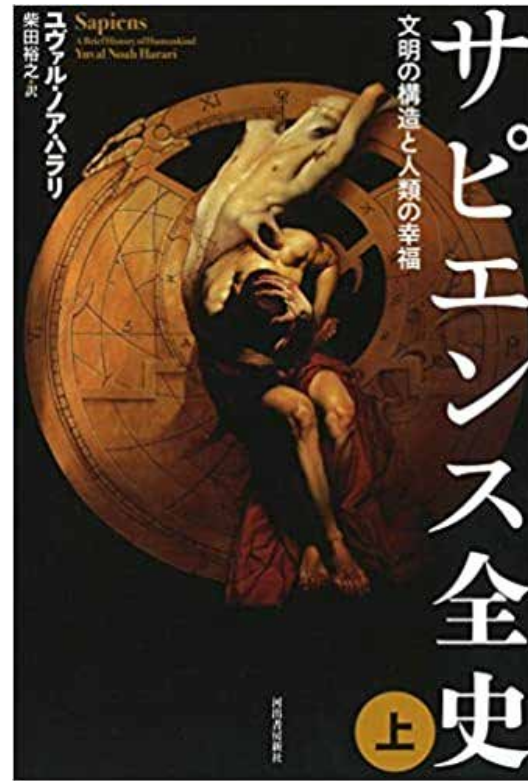
<https://twitter.com/cdixon/status/505118160811728896>

グローバル／ビッグヒストリー

「歴史」の
対象が空間的・社会的
に拡大し、
より分野横
断的になっ
てきた



ジャレド・ダイヤモンド、銃・病原菌・鉄



ユヴァル・ノア・ハラリ、サピエンス全史



デヴィッド・クリスチャン、ビッグヒストリー われわれはどこから来て、どこへ行くのか——宇宙開闢から138億年の「人間」史

European Time Machine



- **欧州**を中心とした4000年の文化遺産デジタル化と活用
- **予算**：10年で1000億円→現在は大幅減額？

「過去のビッグデータ」を探る欧州タイムマシン研究計画, カレントアウェアネス-E, No.389, 2020.04.23, <https://current.ndl.go.jp/e2248/>

Hanyang Time Machine

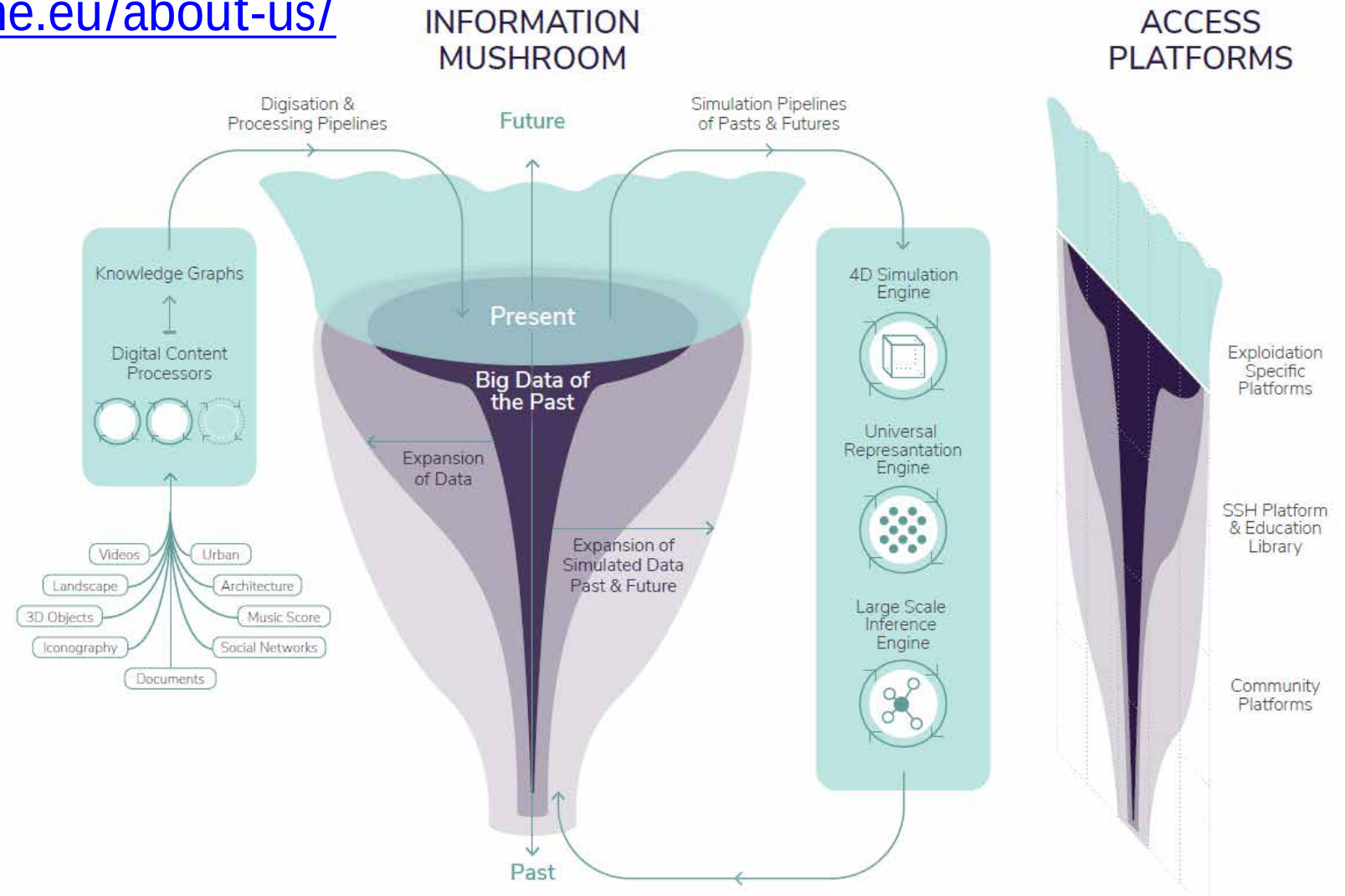


- **韓国・ソウル**の3次元復元（メタバース）と知識ネットワークの構築
- **予算**：7.6百万ドル／年

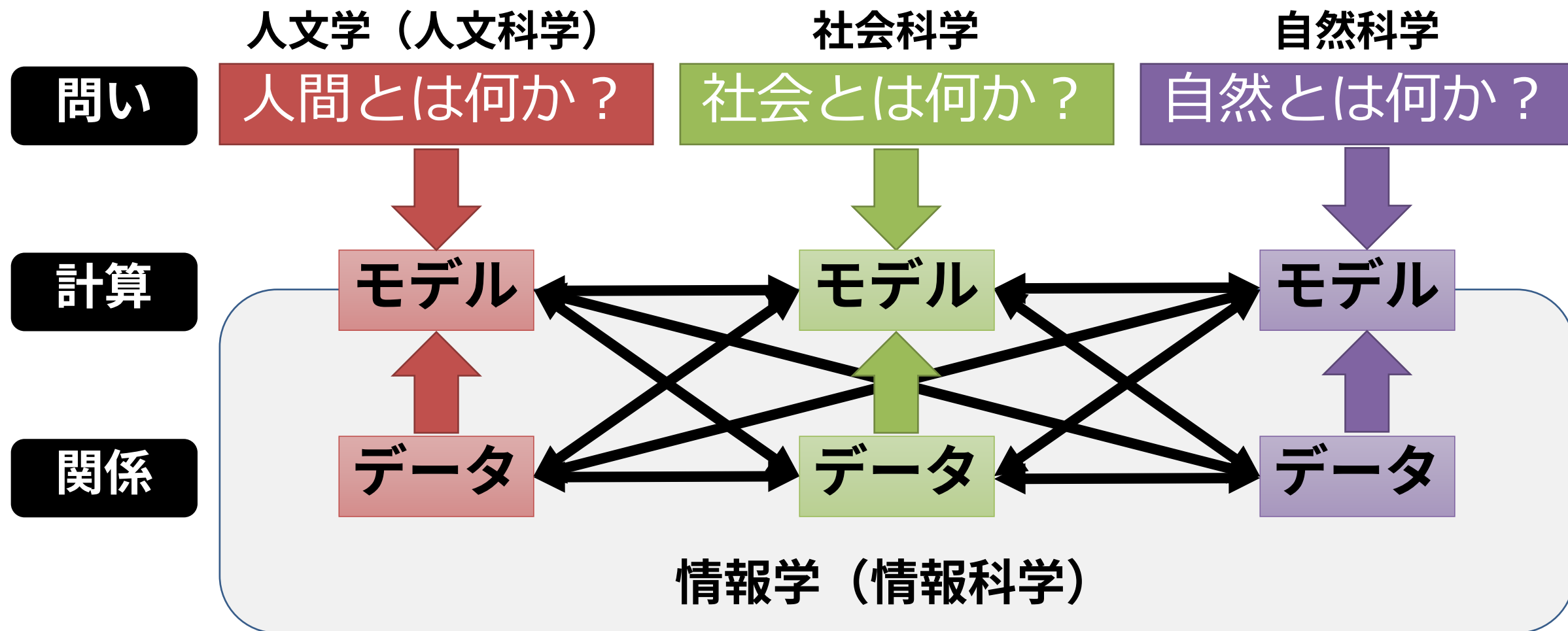
タイムマシンの概念図

<https://www.timemachine.eu/about-us/>

1. **デジタル化**で過去のデータを増やす。
2. **AI**で構造化データを増やす。
3. **シミュレーション**で欠けているデータを補う。
4. 過去を未来に延長する（4D）。



総合知とデジタル研究基盤



謝辞

1. 本発表には、ROIS-DS人文学オープンデータ共同利用センターの特任助教（鈴木 親彦、市野 美夏、カラーヌワット・タリン（当時））の成果を含みます。
2. IIIF Curation Platformの開発には、本間 淳（フェリックス・スタイル）、Tarek Saier（カールスルーエ工科大学）らが大きく貢献しました。
3. 国文学研究資料館や国立国会図書館、人間文化研究機構、国土交通省などが提供するオープンデータは、研究を進める上で必須の存在であり、データ公開に感謝します。
4. 立命館大学には、MapWarperの利用に際して、アドバイスをいただきました。



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