

歴史ビッグデータ：人類が生み出した記録に基づく環境と社会のシームレスな分析



北本 朝展（きたもとあさのぶ）

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

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

ビッグデータとは？

Volume

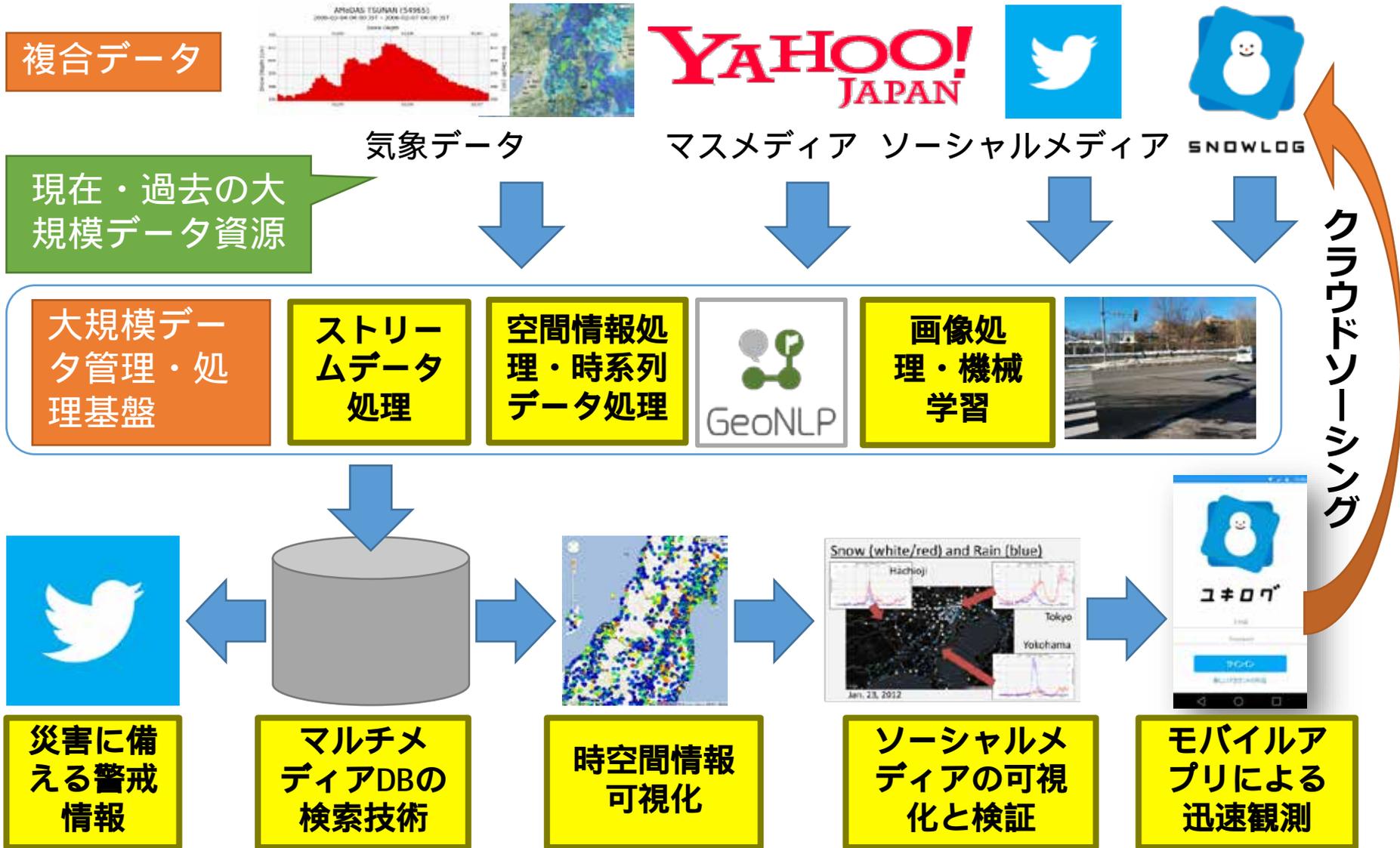
Variety

Velocity

Veracity

- ビッグデータとは、単なる「大きなデータ」ではない。
- データの意外な組み合わせの探求。
- データの目的外利用による新しい発見。
- 歴史ビッグデータという新しい概念の提唱。

ビッグデータシステムの典型的構成



DIASの多様なアプリ群

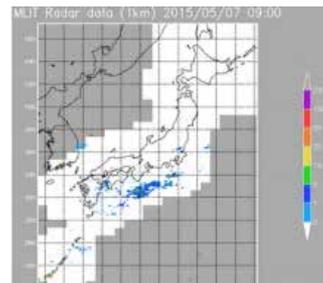
Data Dissemination



Himawari-8 Satellite



Weather Forecast GPV

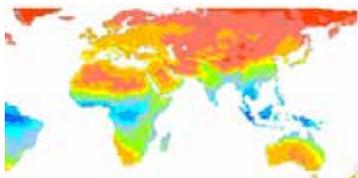


Radar Data

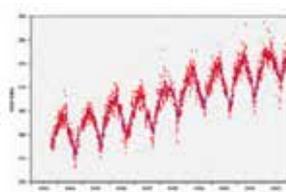


River Telemeters

1. Climate



CMIP5 Model



Visualization Tools

2. Water



Dam Control



Water Management

3. Agriculture



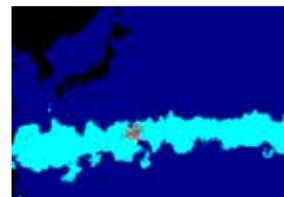
Potential of Rice Crops after Climate Change

2018/7/12

4. Biodiversity



Citizen science-based observations



Fish eggs and growth distribution



Accumulated radar rainfall

デジタル台風

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

- 台風に関するあらゆるデータをアーカイブし、異種のデータを相互連結。
- 過去のアーカイブデータとリアルタイムデータをシームレスに統合。
- ひまわり画像やアメダスデータ40年分など、データの網羅性が大きな売り。
- 利用状況約2000万PV / 年。

100年天気図データベース

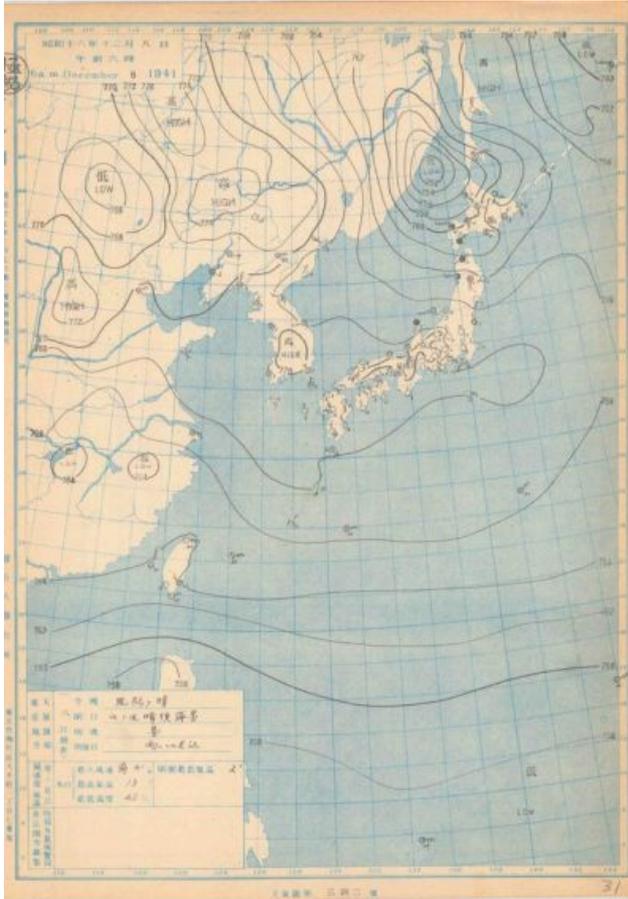
<http://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/>



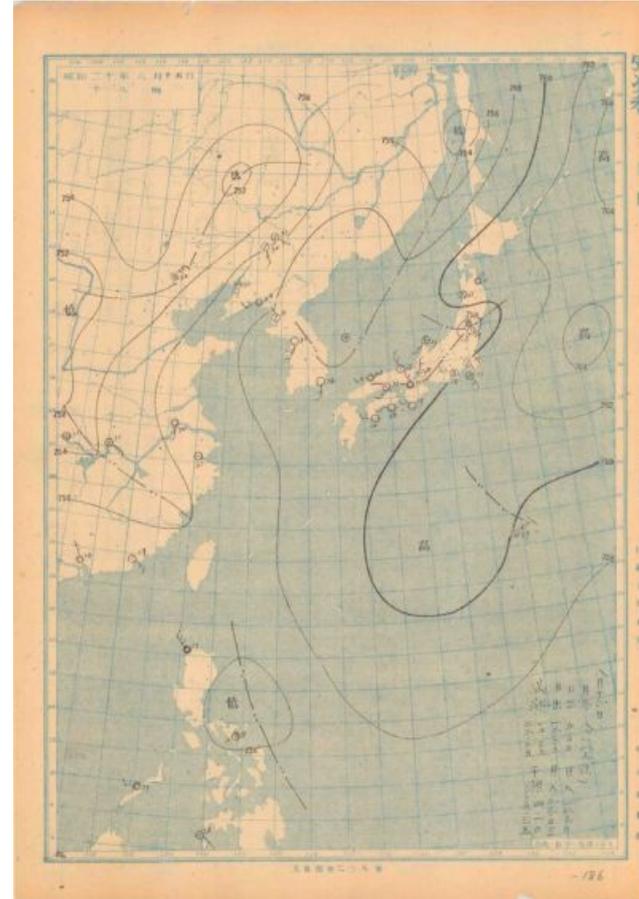
- 1883年3月1日以来の天気図アーカイブ。日時で検索し、過去の天気図にアクセスできる。
- 1883年当時の最新通信技術「電信」が迅速な天気図作成を実現した。
- 気象のみならず社会の動きも反映している。

注：気象庁「天気図」をもとに国立情報学研究所「デジタル台風」が作成

史料としての読み解き



太平洋戦争開戦 (1941年12月8日)



太平洋戦争敗戦 (1945年8月15日)

岡田武松による「観測精神」

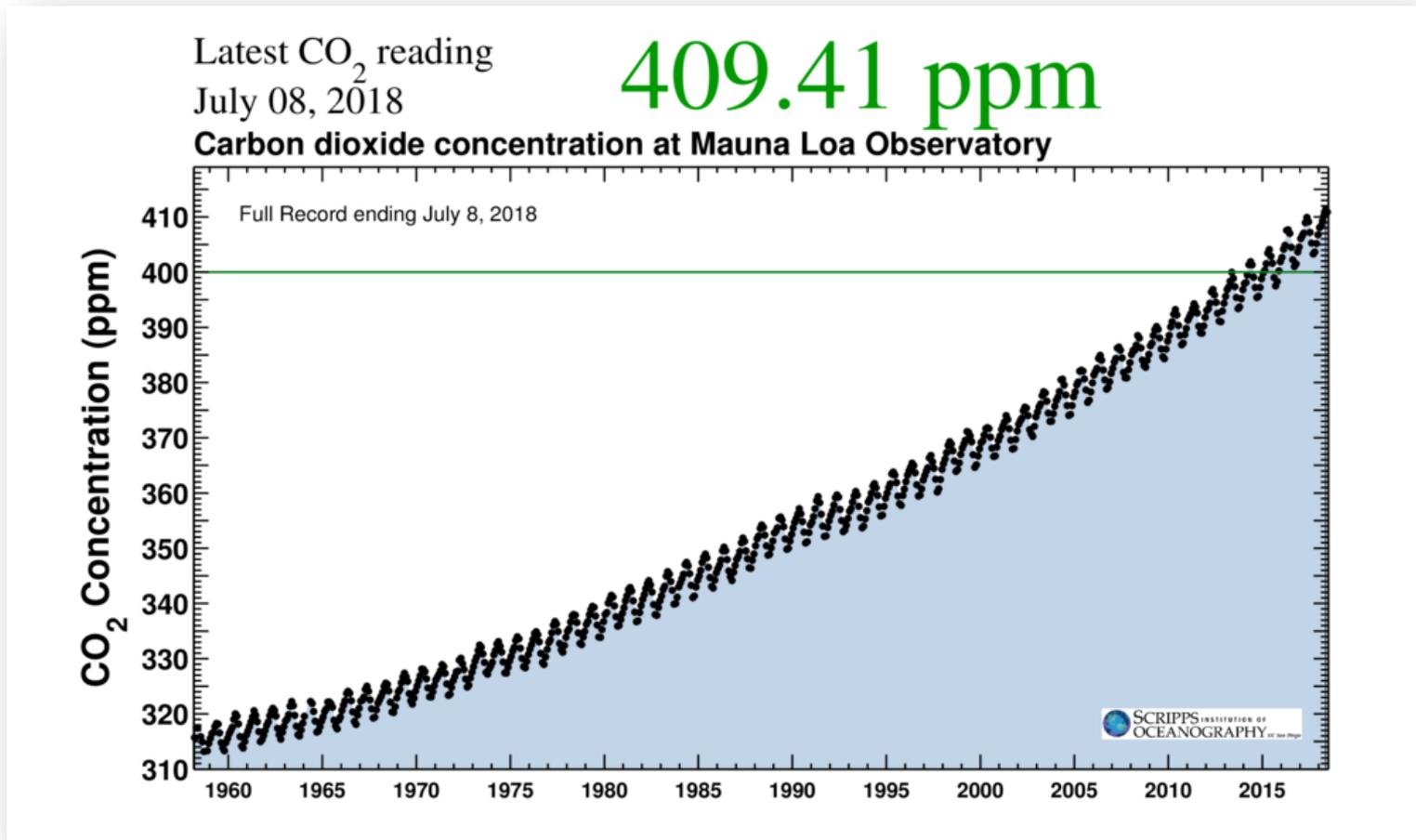
<http://researchmap.jp/jokogkuim-1786/>

観測精神とは、あくまで科学者の精神である。自然現象は二度と繰り返されない。観測とは自然現象を正確に記録することである。同じことが二度と起こらない自然現象を欠測してはいけない。それではデータの価値が激減するからである。まして記録をごまかしたり、好い加減な記録をとったりすることは、科学者として失格である。

(柳田邦男著「空白の天気図」)

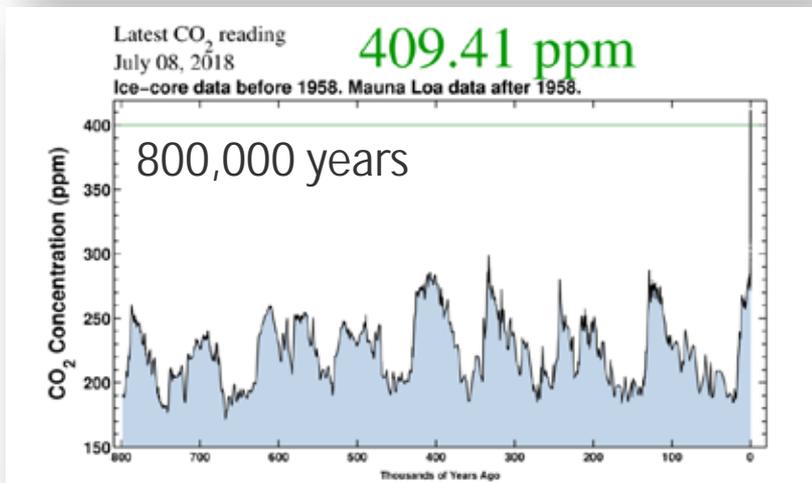
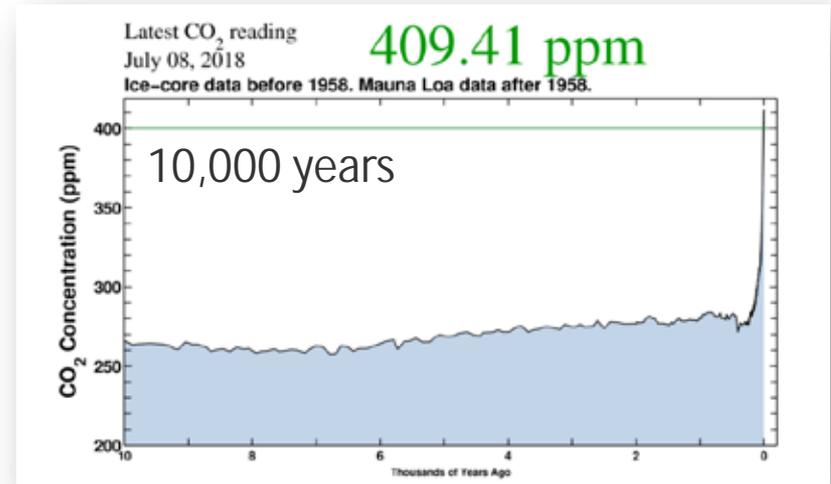
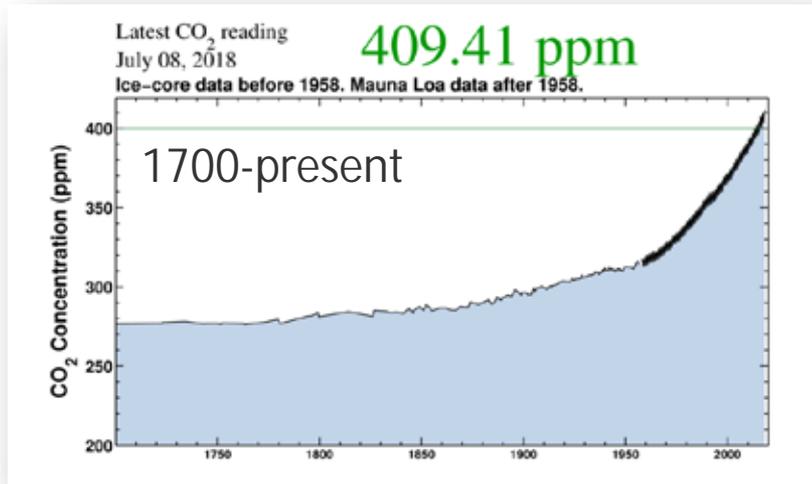
歴史ビッグデータとは？

人間の観測によるデータ



THE KEELING CURVE: <https://scripps.ucsd.edu/programs/keelingcurve/>

証拠から推測したデータ



1958年以降はマウナロアの観測、1958年以前はアイスコアのデータ。

THE KEELING CURVE:

<https://scripps.ucsd.edu/programs/keelingcurve/>

人間による状況記録

状況記録：人間が感覚器（眼など）を用いて世界の状況を観察し、文字などの形式で記録したもの。

1. **現代的状況記録**：ボーンデジタルなデータとして、ソーシャルメディアやスマートフォンなどに残る記録。
2. **歴史的状況記録**：アナログな紙に書かれた文字として、日記などに残る記録。

歴史ビッグデータ

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



歴史的状況記録

現代的状況記録

歴史的観測記録

歴史的環境証拠

歴史データ
としての共通性に着目

状況記録としての共通性に着目

残存する大量の史料の活用



天気



季節



地震

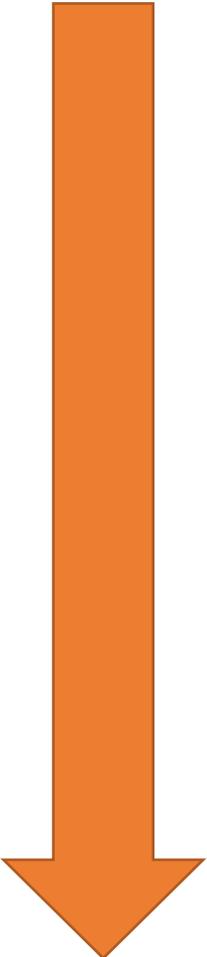


火山噴火

出典：早稲田大学図書館 古典籍総合データベース

1. 江戸時代は、人々の識字率も高く、世界有数の出版大国でもあった。
2. 数億件（？）の史料が残存するが、くずし字を読める人が少なく活用進まず。

構造化ワークフロー

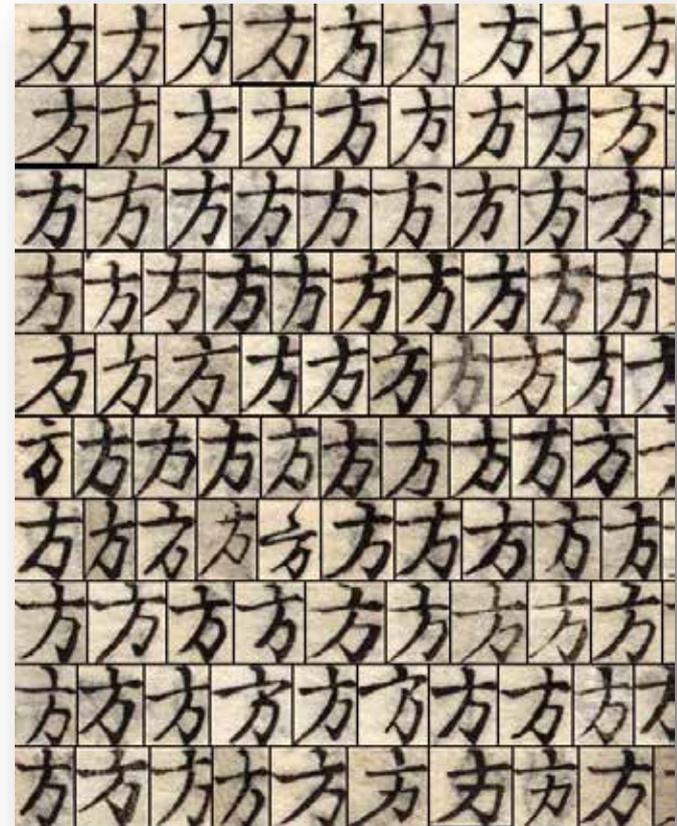
- 
1. 非構造化データ（画像）
 2. 非構造化データ（プレーンテキスト）
 3. 半構造化データ（マークアップテキスト）
 4. 構造化データ（ローデータ）
 5. 構造化データ（解析準備データ）
- 歴史的
状況記録**

非構造化データ：画像→文字



<http://www.honkoku.org/>

人力による電子テキスト化



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

機械による電子テキスト化

歴史的状況記録の形式

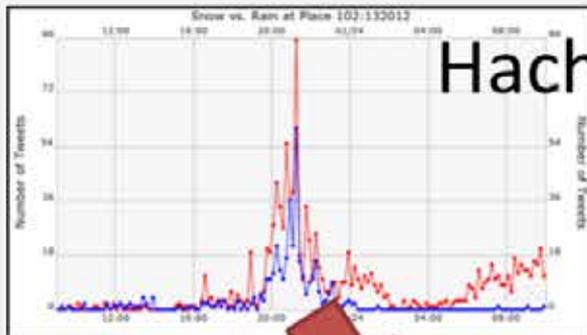
1. **場所**：緯度経度登録、現代地名変換、歴史地名＋地名データベース
2. **時間**：西暦登録、原本表記の保持＋HuTime暦変換サービス
3. **状況**：大分類（災害／現象／経済）、中分類（天気／地震）、自由記述
4. **種類**：原文表記、コード化、数値化
5. **信頼性**：相互比較による専門家評価

日記とツイートの比較

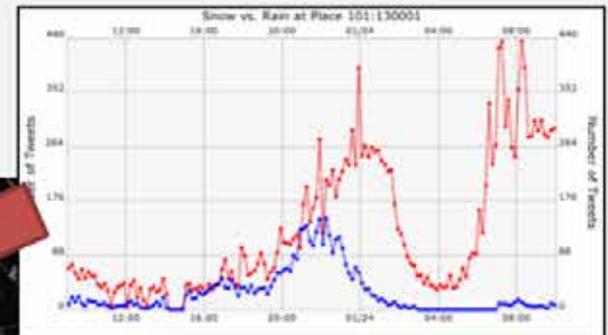


1. 日記とツイートはどこが違うのか？
2. 両者を共通の方法で構造化できるか？
3. 記述の曖昧さなどは、共通の方法（自然言語処理等）で扱えるか？

Snow (white/red) and Rain (blue)

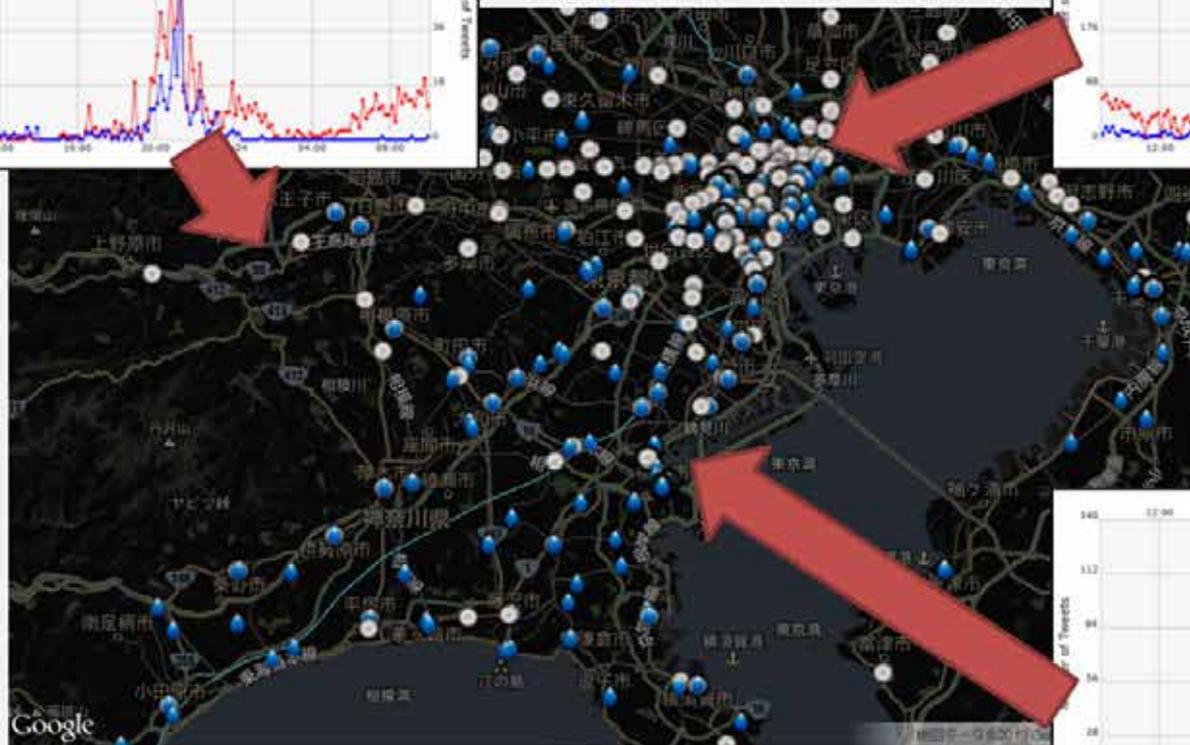


Hachioji



Tokyo

Yokohama



Jan. 23, 2012



ふってきったー: <http://agora.ex.nii.ac.jp/futtekitter/>

「安政江戸台風」の例

帝京大学 平野 淳平氏

CODH 市野 美夏氏

らによる研究成果の紹介

安政江戸台風とは？



安政3年（1856年）8月25日～26日

江戸付近を通過し、**江戸湾周辺**で**高潮**が発生。

浸水、火災で死傷者多数。

『安政風聞集』 卷之中 1856年刊（早稲田大学図書館所蔵）

http://archive.wul.waseda.ac.jp/kosho/wo01/wo01_03959/wo01_03959_0002/wo01_03959_0002_p0011.jpg



#安政江戸台風

1856年9月23日 昼～夜



新島島役所日記 夜
夜入辰巳二相成所々
一円大荒破損多く有
之候

新島島役所日記
辰巳風二成大荒

新島島役所日記
辰巳風天気大雨



#安政江戸台風

1856年9月23日 午後4時



ハリス@下田 16:00

The wind at four P. M. was S. S. E., and continued to haul to S. S. W., at which point the gale was heaviest.

ハリス@下田 16:00

Yesterday at four P. M. the wind began to blow fresh from E. S. E., with rain.



#安政江戸台風

1856年9月23日 夕方～夜中



石川@八王子 夜八ツ時
夜五ツ半時より八ツ
時迄大嵐二而一日大
あれ也、尤上々度八
大あれ也、江戸人死
多し

関口@横浜 夕刻
陰天、夕刻より雨降
出し段々風強く、雷
鳴両三声



#安政江戸台風

1856年9月24日 未明～朝



源兵衛@相馬 暁七ツ半頃
 暁七ツ半頃より止風猛し、大柳式本根返り、水流通り、近年希成大時化、立木八たふれ又南之方程強し

大高@水戸 明方
 明方辰巳風二相成漸く静二相成

三右衛門@玉村町 丑寅
 昨夜大嵐丑寅夜七ツ頃力北風二成る

安政江戸台風の経路推定



風向の変化から、
台風との位置関係
を各地で推定。

各地の推定を集約
して、台風の経路
を推定。

東京湾高潮の実例
として、今後の災害
に関する教訓を得る。

歴史ビッグデータへの展開

1. 古気候の長期再解析や、気候変動の影響評価などへの利用。
2. ビッグデータの「目的外利用」（プローブデータ→通れた道マップ）。
3. 地震（1855年安政江戸地震）と台風とを、複合災害として分析できるか？
4. 江戸の人口分布と台風被害、経済成長などの関係を統合的に分析できるか？

シームレスなアーカイブに向けて

シームレスなアーカイブ

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

The screenshot shows the 'Digital Typhoon' website interface. At the top, there's a navigation bar with 'Home', 'Earth', and 'Digital Typhoon'. Below that, a header section reads 'デジタル台風：台風画像と台風情報'. A main content area features a 'リアルタイム台風速報' (Real-time Typhoon Report) for Typhoon 201724 (WVP). It includes a satellite image of the typhoon and a list of related news items. On the right, there's a '台風発生数' (Typhoon Occurrence Count) section with statistics for the current year and historical data. At the bottom, there's a '台風データベース' (Typhoon Database) section with a list of search criteria, and a '静止気象衛星画像 (地球)' (Static Meteorological Satellite Image (Earth)) showing a satellite view of the Earth with a typhoon visible over the Pacific Ocean.

- **他の気象データベースとの最大の違い**：過去と現在を分離しないというコンセプトに基づき構築。
- **歴史性を重視**：現在データと過去データをシームレスに接続。
- **統合性を重視**：環境データと社会データをシームレスに接続。

シームレス：過去と現在

1. **現在とは最新の過去**であり、両者は時間という連続体を共有している。
2. **現在の技術**を、過去のデータに適用するため、ロバスト性を強化する。
3. **過去のデータを現在のプラットフォームにシフト**し、現代の人々が活用する。
4. 現在と過去の共通性に着目し、**長期・連続的なビッグデータ**を生成する。

シームレス：環境と社会

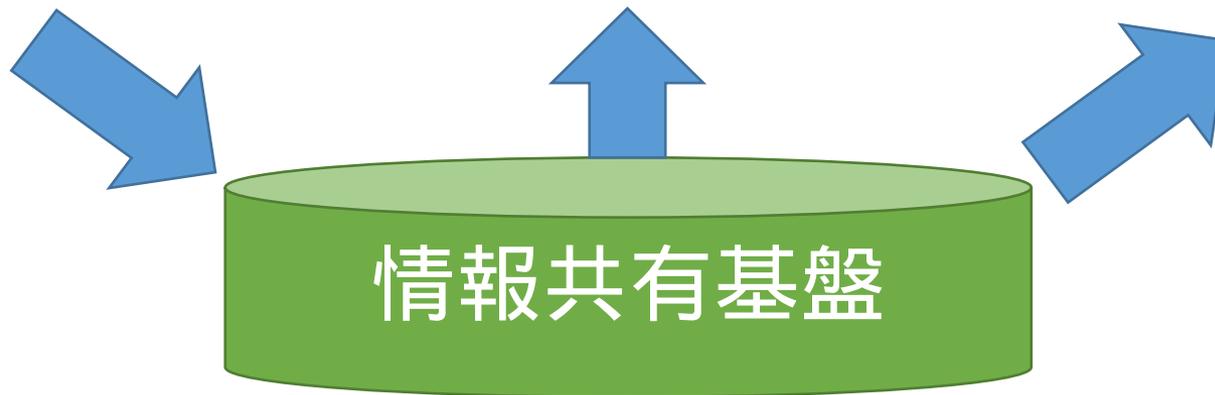
1. 人間の記録は、環境を見る精神のあらわれであり、一種の測定でもある。
2. 情報学の知見を加えることで、扱いづらいデータを扱いやすく変換する。
3. 史料をエビデンスとしてアクセス可能とし、データを検証可能とする。
4. 客観的な環境証拠に人間の状況記録を加え、社会のビッグデータを生成する。

情報共有基盤の構築

状況記録

検索・可視化

APIアクセス



1. 歴史的状況記録のスキーマを定め、分野横断的に蓄積する機能を実現する。
2. 検索・可視化だけでなく、API経由のオープンアクセスによる研究利用を促進する。

歴史ビッグデータへの誘い

- 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～**」などの会議を開催。
- くずし字翻刻や史料所在情報共有のための**参加型プラットフォーム**を構築。
- **人文学分野にもデータ共有文化を普及**させ、分野横断的なデータ共有を推進。
- 興味がありましたら、ぜひご連絡を！

参考リンク

- **人文学オープンデータ共同利用センター**
 - <http://codh.rois.ac.jp/>
- **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～**
 - <http://codh.rois.ac.jp/seminar/historical-big-data-20180312/>
- **デジタル台風**
 - <http://agora.ex.nii.ac.jp/digital-typhoon/>
- **Researchmap**
 - <http://researchmap.jp/kitamoto/>