



歴史ビッグデータと 歴史的状況記録

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（CODH）

国立情報学研究所

北本 朝展、市野 美夏

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

ビッグデータとは？

Volume

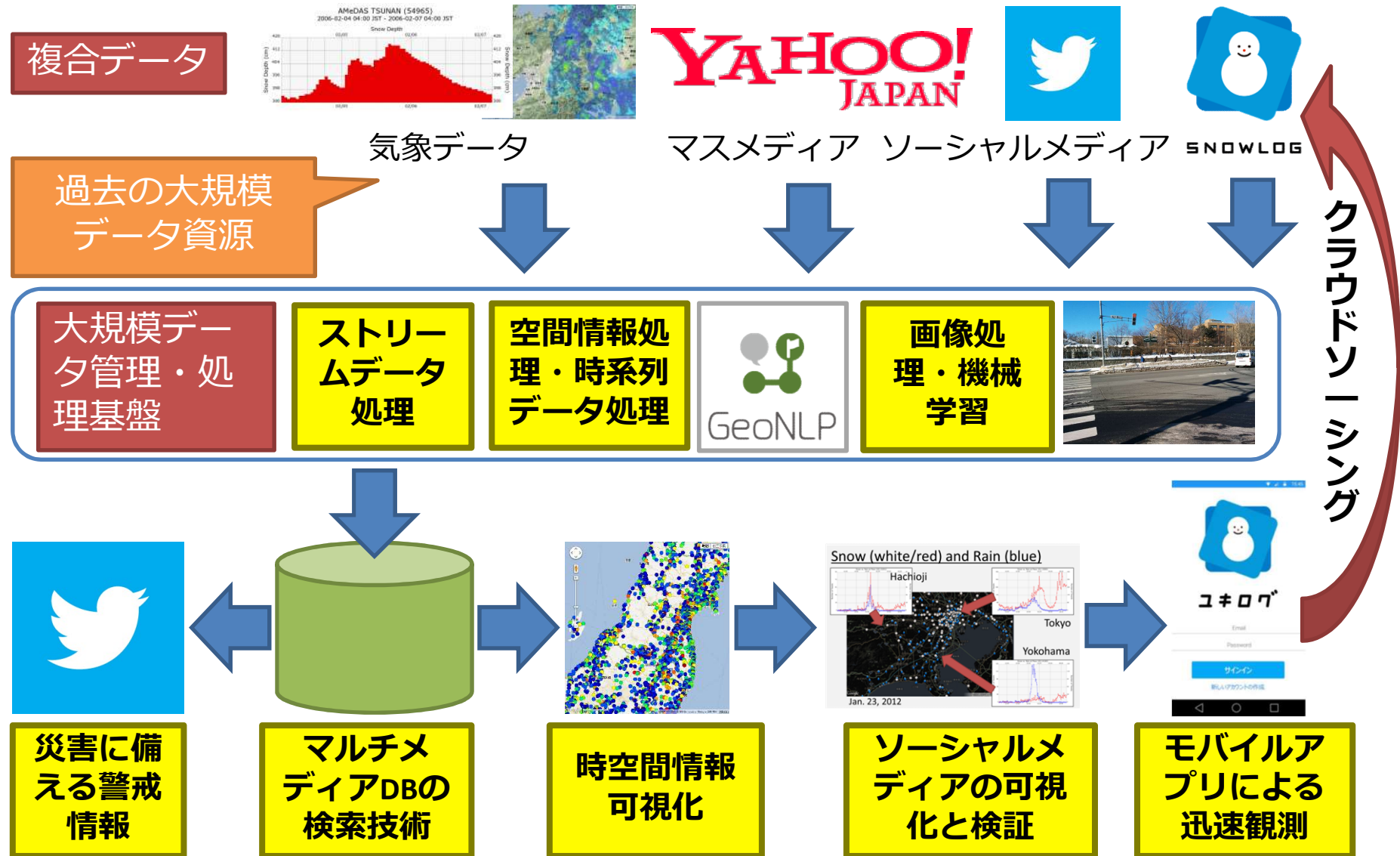
Variety

Velocity

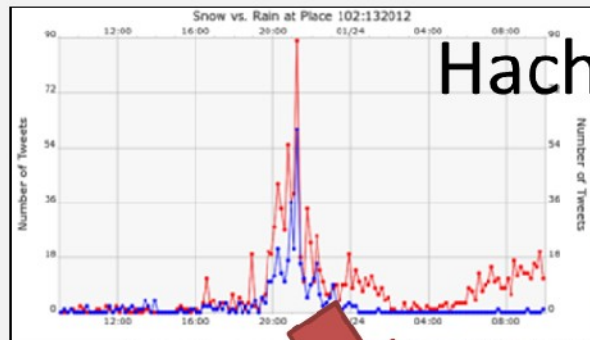
Veracity

- 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。

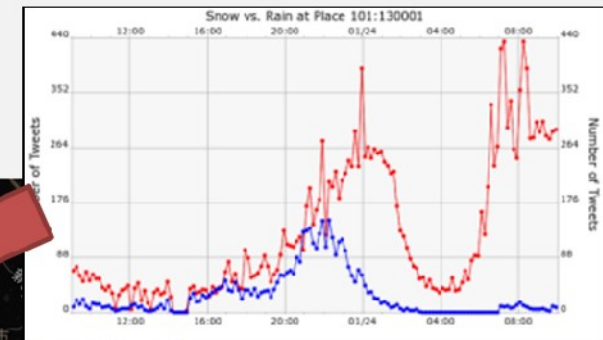
ビッグデータシステムの典型的構成



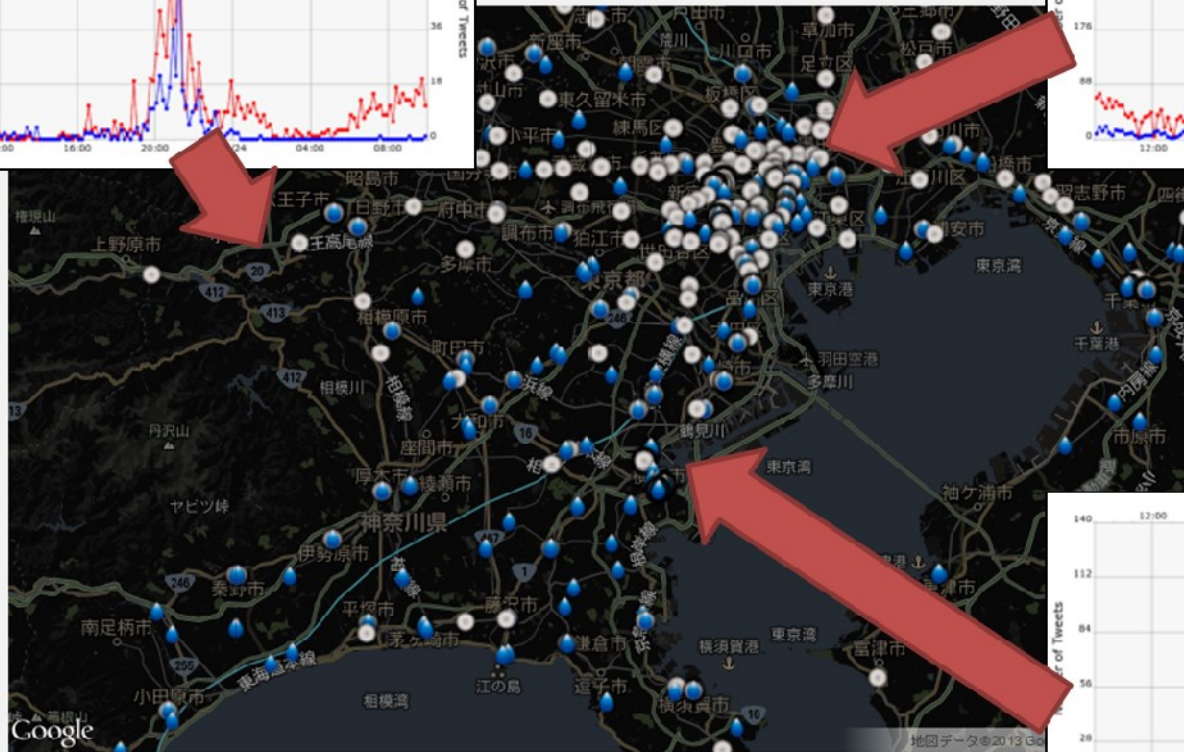
Snow (white/red) and Rain (blue)



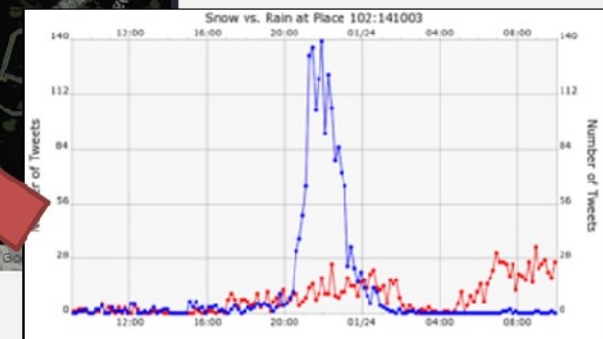
Hachioji



Tokyo



Yokohama



Jan. 23, 2012

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

状況記録とは？

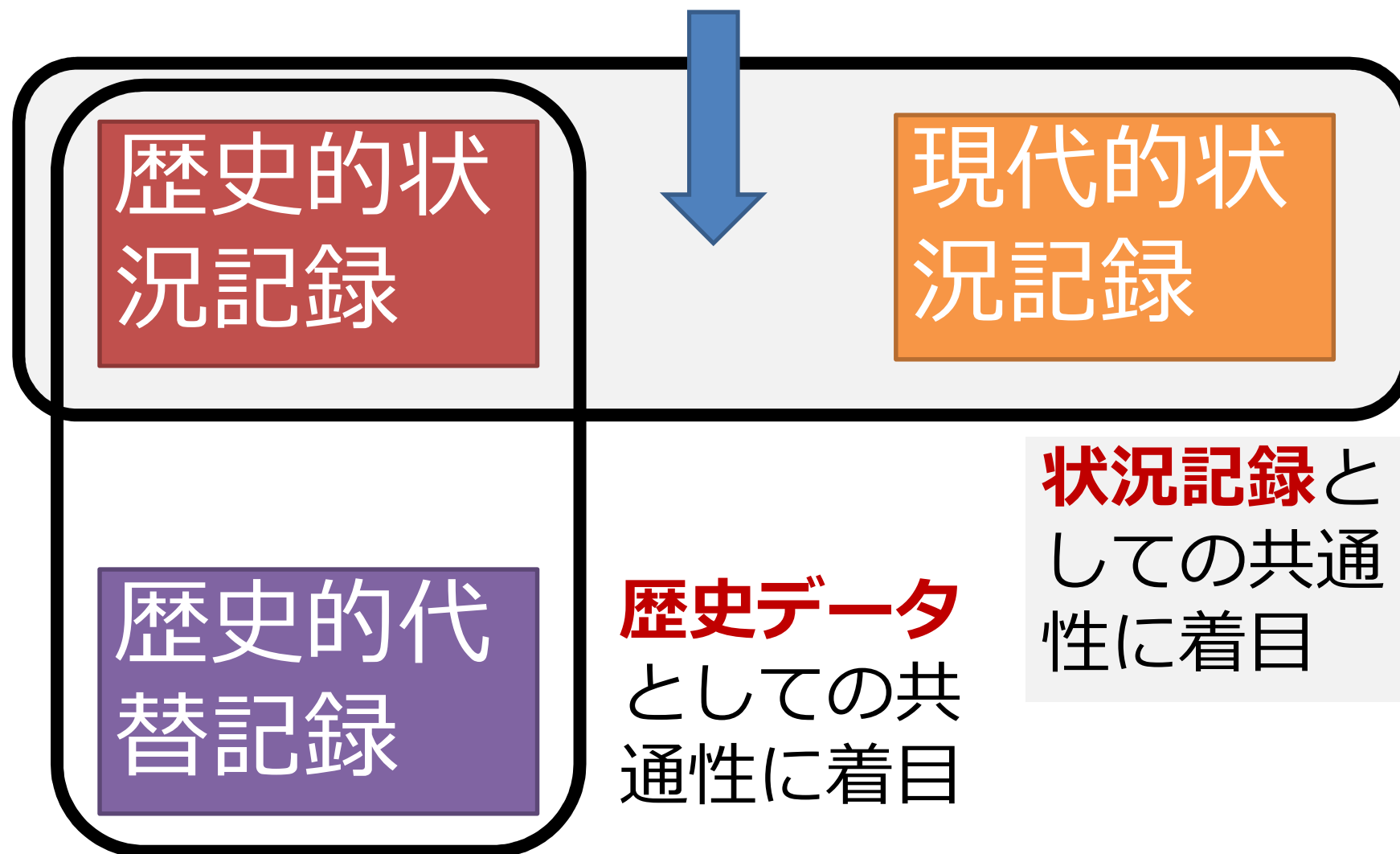
状況記録：人間が感覚器（眼など）を用いて世界の状況を観察し、文字などの形式で記録したもの。

- 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。

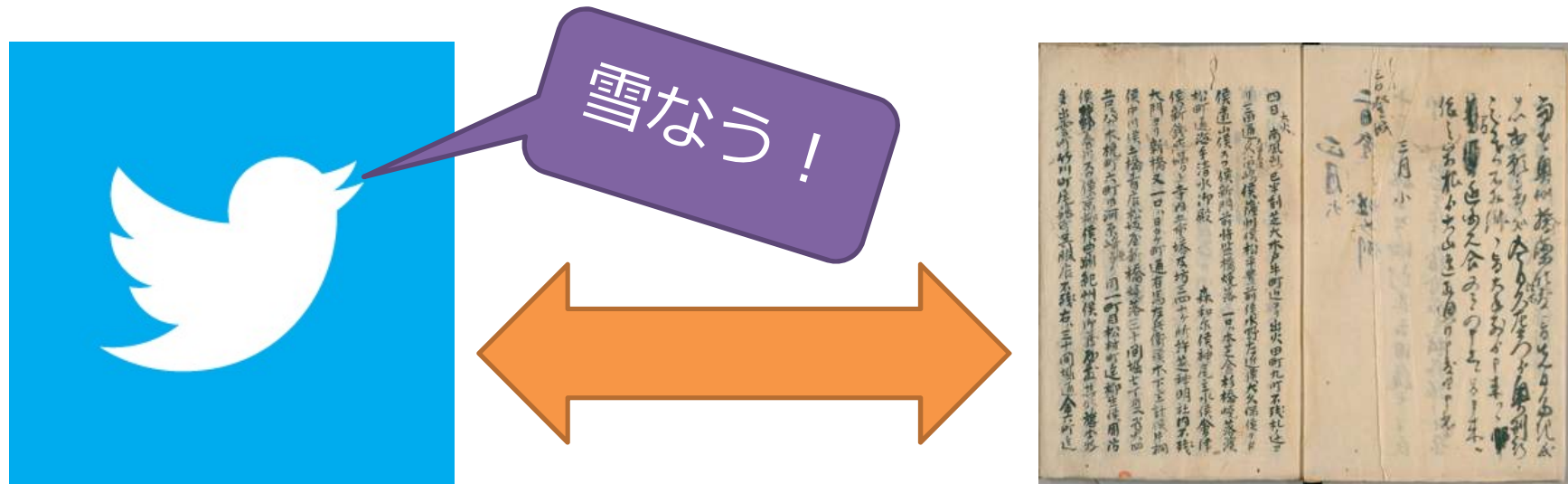
過去の痕跡の調査

1. **歴史的代替記録**：年輪やアイスコア等の自然物を対象に、酸素同位体比などを調査し、**客観的なエビデンス**を得る。
 2. **歴史的状況記録**：日記などの人工物を対象に、記述内容などを調査し、**主観的なエビデンス**を得る。
- ただし、**高い品質で客観性や信頼性を満たす歴史的状況記録**も存在する。

歴史ビッグデータ



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？

歴史ビッグデータの課題

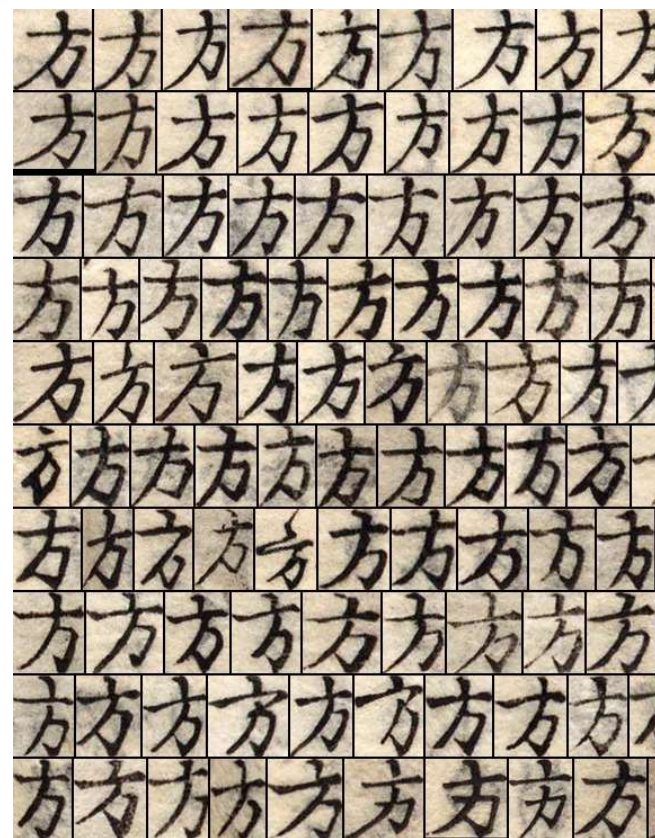
1. 現代的状況記録で発展した技術体系を歴史的状況記録に取り込む。
2. 歴史的状況記録に特有の技術が必要な場合は新たに開発する。
3. 分野依存性を考慮した上で、互換性や再利用性を最大化する。
4. 情報共有のための基盤を構築し、多くの人が共通的に利用可能とする。

ボーンアナログの課題



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

人力による電子テキスト化



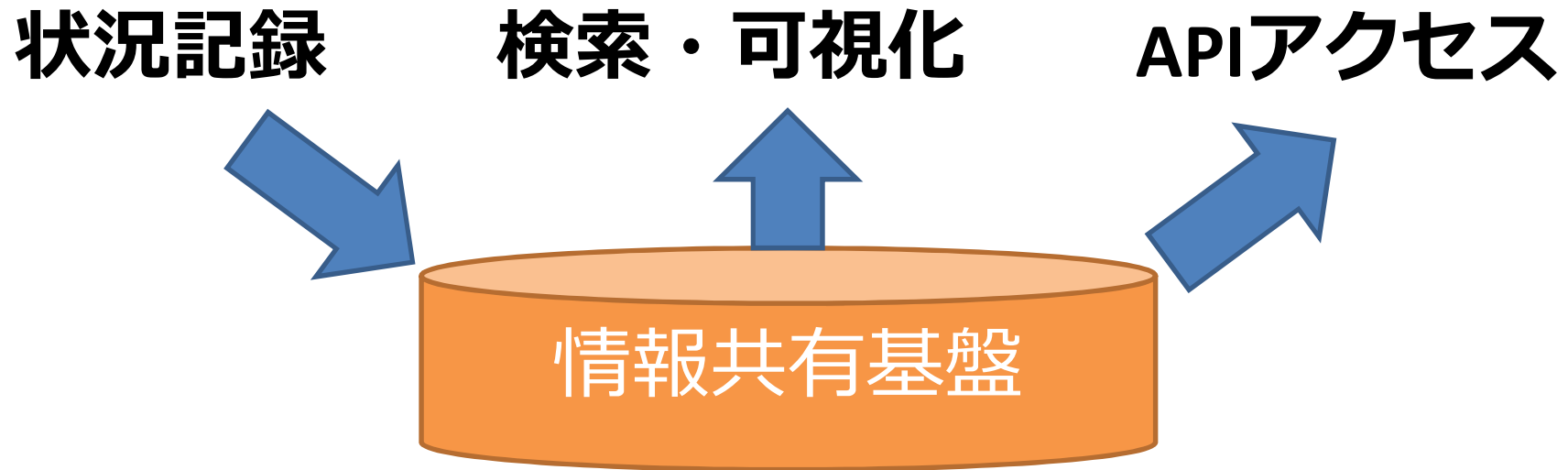
<http://codh.rois.ac.jp/char-shape/>

機械による電子テキスト化

状況記録スキーマ

1. **場所**：緯度・経度、現代地名、歴史地名
＋地名データベース
2. **時間**：登録時に西暦変換、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）、状況の詳細記述
4. **種類**：数値、テキスト（曖昧表現）
5. **信頼性**：相互参照・比較による評価

情報共有基盤の構築



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。

具体的な研究課題

1. **情報基盤**：歴史的状況記録スキーマ検討、データベース構築、API提供。
 2. **データ構造化**：資料発掘、構造化データ入力・テキスト化（人間／機械）
 3. **データ分析**：信頼性評価、補間（同化）による2/3/4次元化、新たな知識の発見
- 2が入力側、3が出力側に相当する。
 - 双方向的に成果を共有することも重要。u

資料情報共有ツール

どの資料にどんな記述があるか？ **本文レベルでの所在情報の共有**から始める。

歴史的状況記録

Browse Collections Browse Items

SEARCH (28 TOTAL)

Query: Query type: Exact match Record types: item, file, collection 1 of 3

Record Type	Title
Collection	 あきる野市教育委員会 『儀三郎日記』 明治の元締1～4
Collection	 東北アジア研究センター 叢書 第38号 佐藤大介編著 『18～19世紀

人文学データと共有文化

1. 「歴史的状況記録」に関連する研究を進める研究者は、全国に散在している。
2. お互いに成果を共有することが難しく、活動はバラバラに進んでいる。
3. デジタル技術を活用した基盤を構築して、協力できる部分を増やせないか？
4. 人文学データに共有の文化を普及させるための課題を少しずつ解決する。



歴史ビッグデータと 歴史的状況記録

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（CODH）

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