

読み解きのかたち オフサイト地域研 究に向けて

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
北本 朝展 (KITAMOTO Asanobu)

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

市民による宇宙考古学の顛末



- 15歳少年が衛星画像からマヤ遺跡を発見したと主張。
- 専門家が「遺跡ではない」と否定。
- ジングル中の人工的形状 = 遺跡とは限らない。

<http://natgeo.nikkeibp.co.jp/atcl/news/16/051200164/>

アームチェア・スカラー

- **アームチェア・スカラー**：「書齋」のアームチェアに座りながら研究すること。
- **批判的なニュアンス**でしばしば言及される。フィールドに行かずに何がわかる？
 1. フィールド（現地）に行くだけでは得られない知識は何か？
 2. ビッグデータやツールの発展によって、新しく可能になったことは何か？

オンサイト / オフサイト

- **災害対応**では、オンサイト / オフサイトという区別がなされる。
- **オンサイト**：現地にいること。現地における様々なアクションを担当する。
- **オフサイト**：遠隔地にいること。現地から集約した情報の意思決定を担当する。
- **データ駆動型の地域研究**は、主にオフサイトで進められる地域研究？

コンテンツ / コンテクト

- **データ**に関するコンテンツ / コンテクトの区別は様々な定義がある。
- **コンテンツ**：データの内部で完結する解釈と意味。
- **コンテクト**：データの外部の世界との関係性から生まれる解釈と意味。
- **機械（人工知能）**はコンテンツの解釈に向く。**人間**の強みはコンテクトの解釈。

ビッグデータのインパクト

1. **Google Maps / Google Earth** : 世界中の地表が、上空・地上から調査可能（ただし更新頻度や精度の問題あり）。
2. **検索エンジン** : 世界中の情報が、キーワードで調査可能（ただしキーワード設定やディープウェブの問題あり）。
3. **ソーシャルメディア** : 世界中のローカル情報を、人々が発信（ただし地域や内容ごとのバイアスの問題あり）。

機械によるコンテンツ解析

1. **広域性**：人間が移動できない広い領域を調べられる。
2. **大規模性**：人間が観察できない量のデータも処理できる。
3. **迅速性**：情報収集経路が整備されていれば、リアルタイムで把握できる。
4. **定量性**：時空を超えて、基準を揃えたりバイアスを補正したりできる。

コンテクストの重要性



<http://researchmap.jp/jo8rg8gdj-1786/>

- マヤ遺跡を「発見」と信じたのは、ジャングルに人工的な地物があったから。
- 我々も、シルクロード遺跡調査で、同様の勘違いをしそうになったことがある。
- 様々な証拠を組み合わせた「批判的検討」をするには、証拠を取り巻くコンテクストの理解が必要。

コンテンツ + コンテクト

- **コンテンツ**は（定義により）場に依存しないので、オフサイトでの分析も可能。
- **コンテクト**は、オンサイト（現地）とオフサイト（現地外）に分離できる。
- **オンサイト・コンテクト**は、「現地」で養う必要性が高い。
- **オフサイト・コンテクト**は、現地でなくても学ぶことが可能では？

関連しそうなケーススタディ

1. **GeoNLP** : 大規模テキストデータを自動的にコンテンツ解析・可視化する例。
2. **シルクロード遺跡** : 複数データのエビデンスを人間が批判的に解釈する例。
3. **エレクトリカル・ジャパン** : ビッグデータを活用して人間が解釈する例。
4. **メモリーグラフ** : 専門家ではない市民でも空間的な解釈に貢献できる例。

クライシスと地域情報

いつ？

どこで？

何が？

分析：

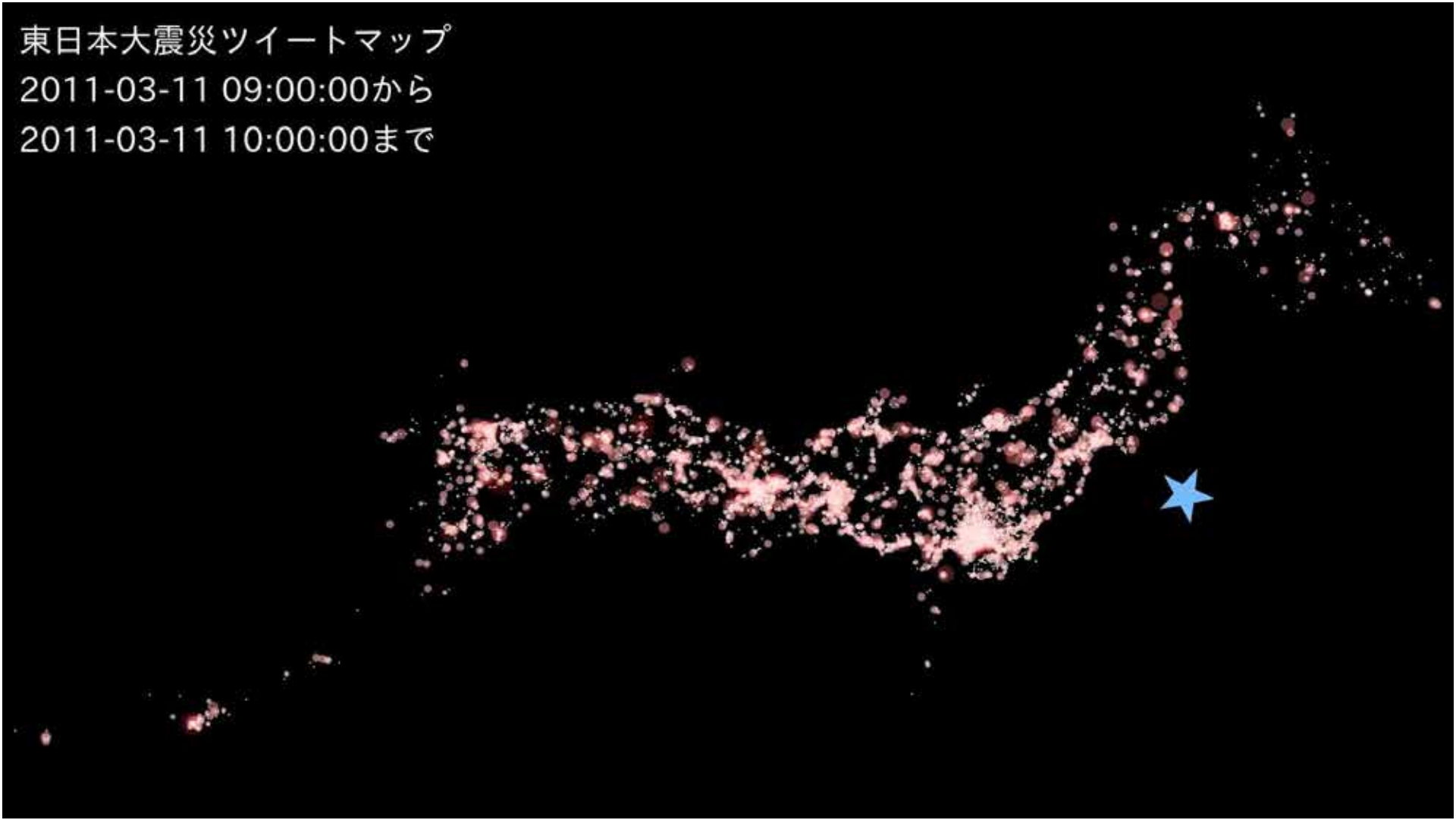
どこが言及されているか？



創生：

その情報をどう見せるか？

東日本大震災ツイートマップ
2011-03-11 09:00:00から
2011-03-11 10:00:00まで



GeoNLPによるジオタギング

<http://agora.ex.nii.ac.jp/GeoNLP/>

Q. 何が難しいの？ A. 地名の表記が曖昧！

「第一小に避難した」 → 全国に「第一小学校」は269校もある！

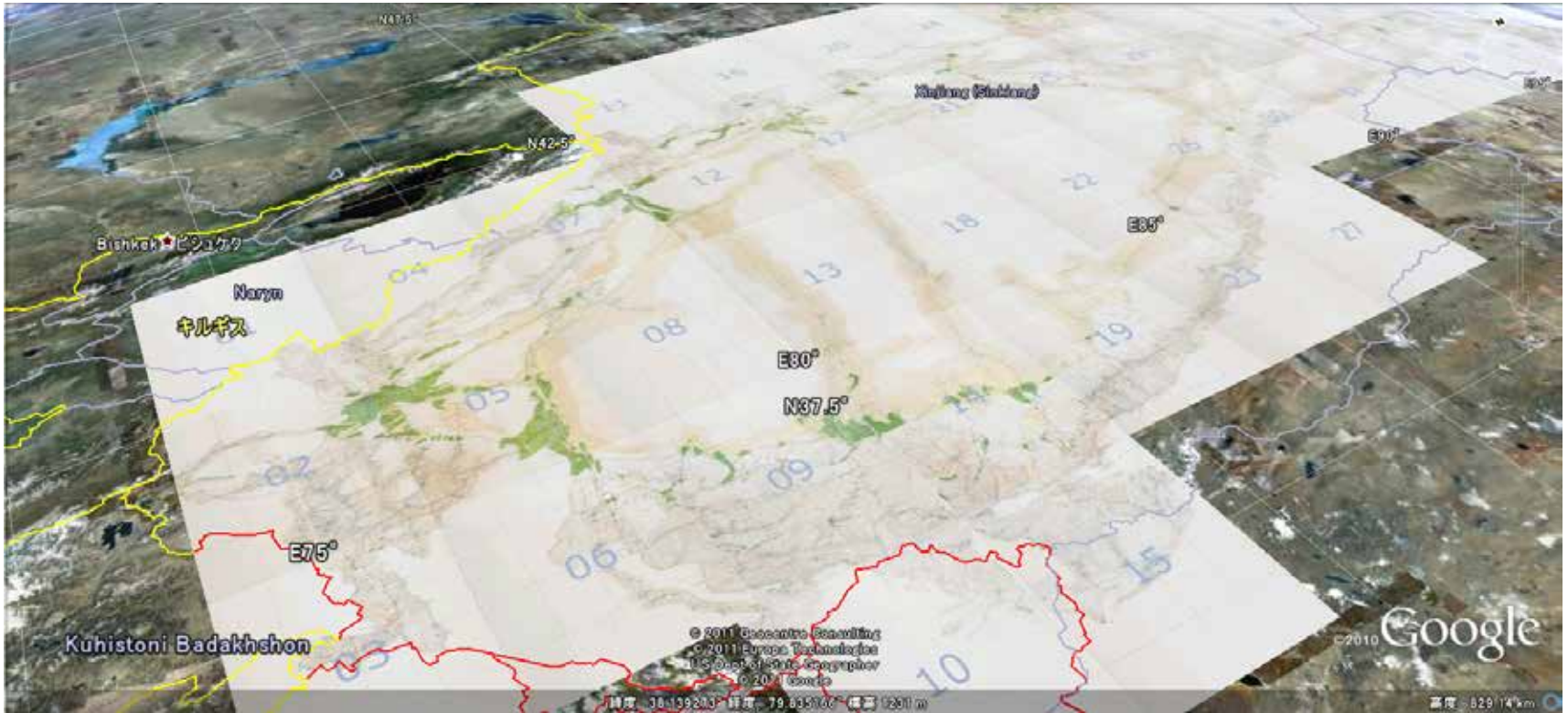
全ツイート 179,286,297 → 地名語含むツイート21,398,123 (12%) を抽出 (地名語出現数としては32,956,427) (参考: ジオタグつき0.15%)

地名語辞書: 約62万 (国土交通省、SaveMLAKなど) + アドレスマッチング

処理速度: 4万ツイート / 1コア・時 → 112コアで約40時間かけて処理

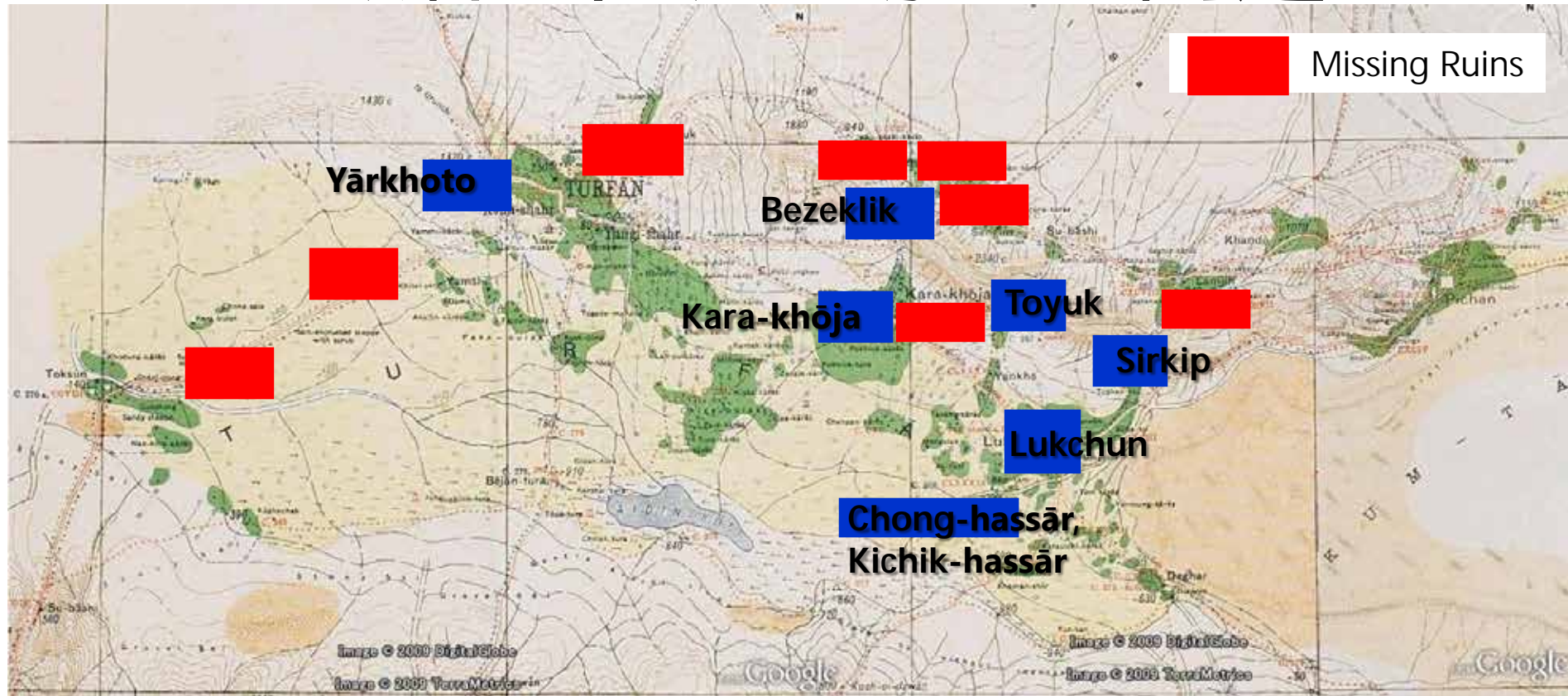
大規模なデータセットを対象とした迅速なマッピングは、コンテンツ解析の典型例。

シルクロード遺跡



地図・写真の読み解きから「見失った」遺跡を再発見

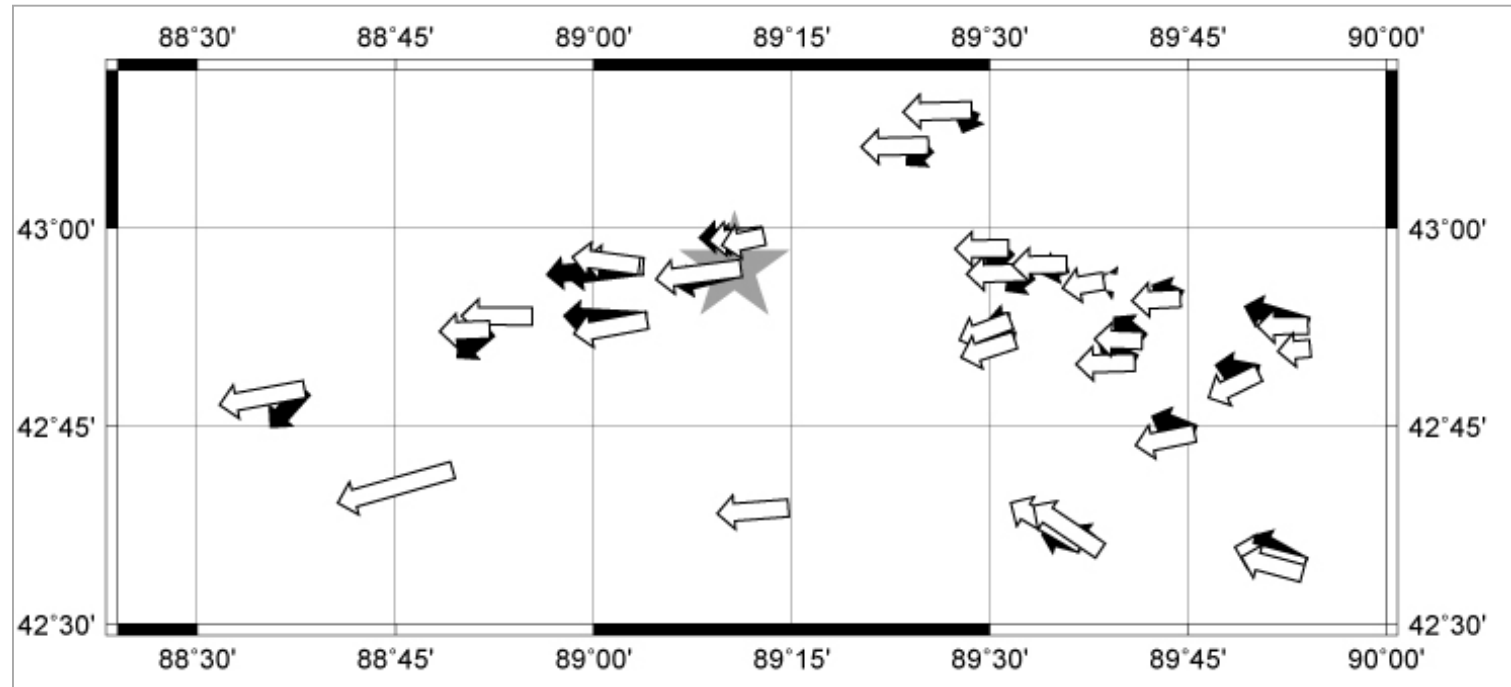
「所在不明遺跡」の問題



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

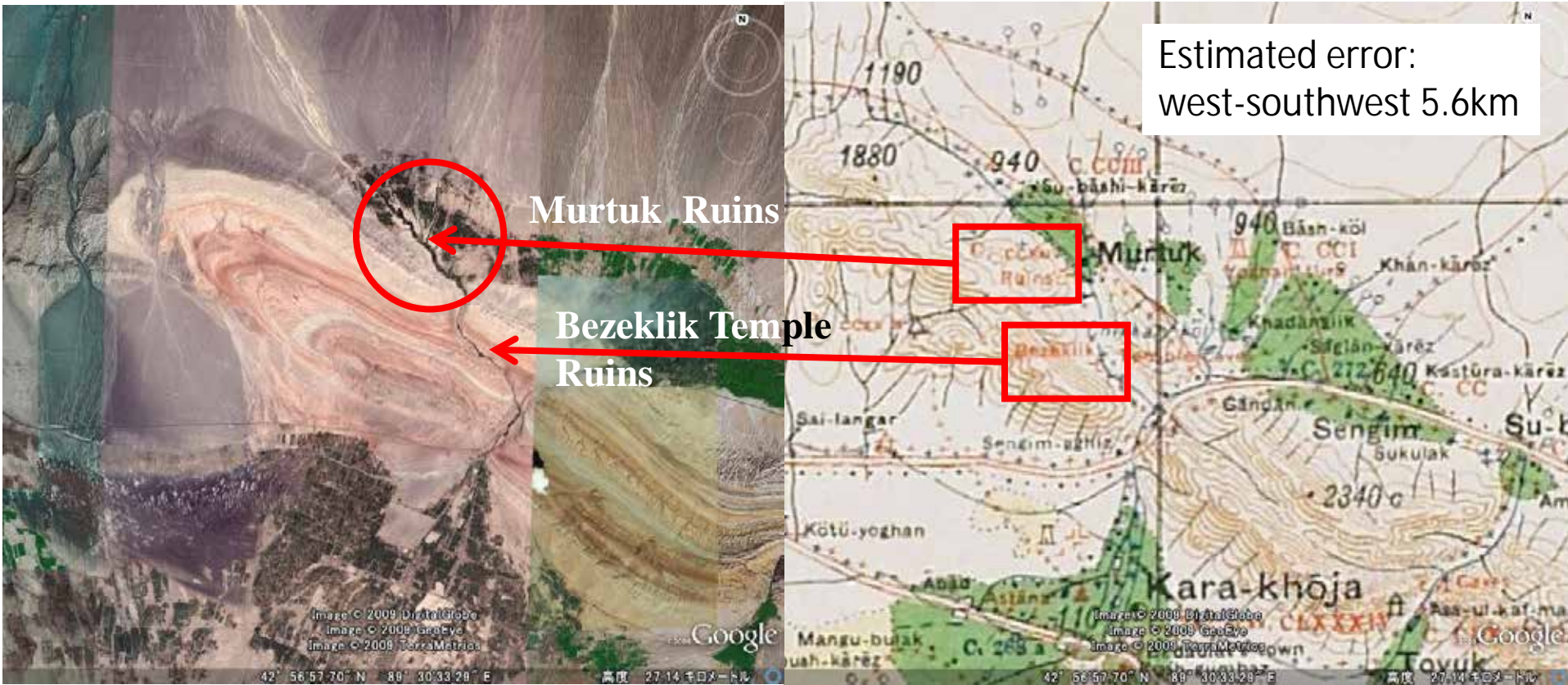
トルファンの古地図誤差分布



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

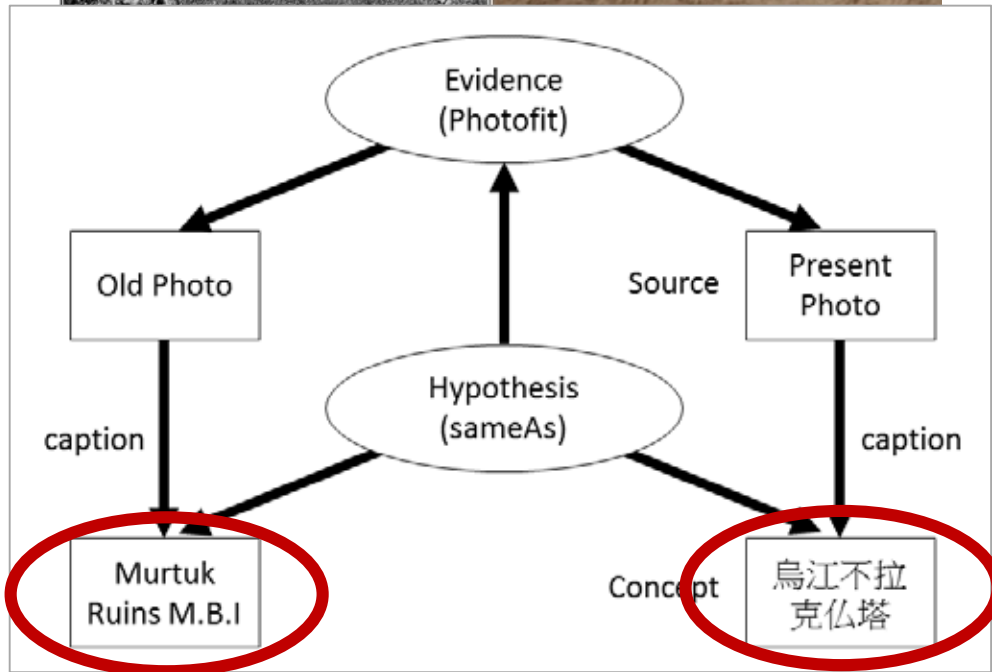
経度方向に誤差が大きいのは、20世紀初頭
は経度測定が技術的に難しかったため。

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



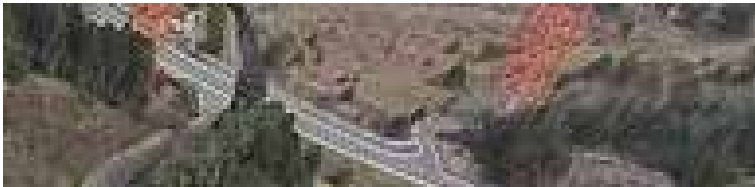
- 複数のデータから得られるエビデンスを批判的に検証し、別名の遺跡を統合できた。
- 背景知識（コンテキスト）が必要となるため、機械による自動化は困難。



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 (2017年4月現在7,000件以上)
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

マッピングのための写真解釈



- **特徴的な人工物**（道路、送電線、施設）を使って場所を絞り込み。
- **特徴的な地形**を読み解き、実際の地図と照合して位置決め。
- **地域に関する知識は、解釈の質を改善する。**



メモリーグラフ

<http://codh.rois.ac.jp/memory-platform/>

- **メモリーグラフ** = 記憶を重ねる新しい写真術。
- カメラのファインダー上に過去の写真を半透明で重ね、現在の風景と構図を合わせてシャッターを押す。
- 古写真を対象とした**文化遺産のフィールドワーク**、**災害前後の定点観測写真**や、**観光地を巡るコンテンツツーリズム**など。
- 旧地域研の共同研究費等からも支援。

メモリーグラフの例



2016/07/23 -> 2017/04/28

コンテンツ → コンテクスト

- **機械（人工知能）**も活用しつつ、コンテンツ解析はコンテクストなしでも可能。
- **コンテクストに基づく解釈**は、背景知識がある人ほど深まる。
- **オフサイトも一つの「現場」**であり、そこで生まれる背景知識がある。
- **オンサイトにおけるコンテクスト**だけに頼ると、解釈の可能性が狭まるのでは？



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行う。**
 2. 人文学のデータを用いて情報学・統計学の研究を行う。

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

研究者・機
械・市民の共
創トライアド

広げる

オフサイト地域研究に向けて

- 「**現地**」に行かないと体感できないコンテキストは確かにある。
- **コンテンツ**解析については、機械との協調が今後はますます重要になる。
- オフサイトの「**現場**」で大量データを見ないと体感できないコンテキストもある。
- **オンサイトとオフサイトを合わせた解釈**は、地域研究に新たな可能性をもたらす。

関連ウェブサイト

- 人文学オープンデータ共同利用センター
 - <http://codh.rois.ac.jp/>
- デジタル・シルクロード
 - <http://dsr.nii.ac.jp/>
- エレクトリカル・ジャパン
 - <http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/energy/electrical-japan/>
- 東日本大震災ビッグデータ分析
 - <http://digital-typhoon.tv/prj311/>