

古典籍の画像分析に向けて

北本 朝展（きたもと あさのぶ）

国立情報学研究所 / 総合研究大学院大学

<http://agora.ex.nii.ac.jp/~kitamoto/>

データベースを作るだけでは足りない！

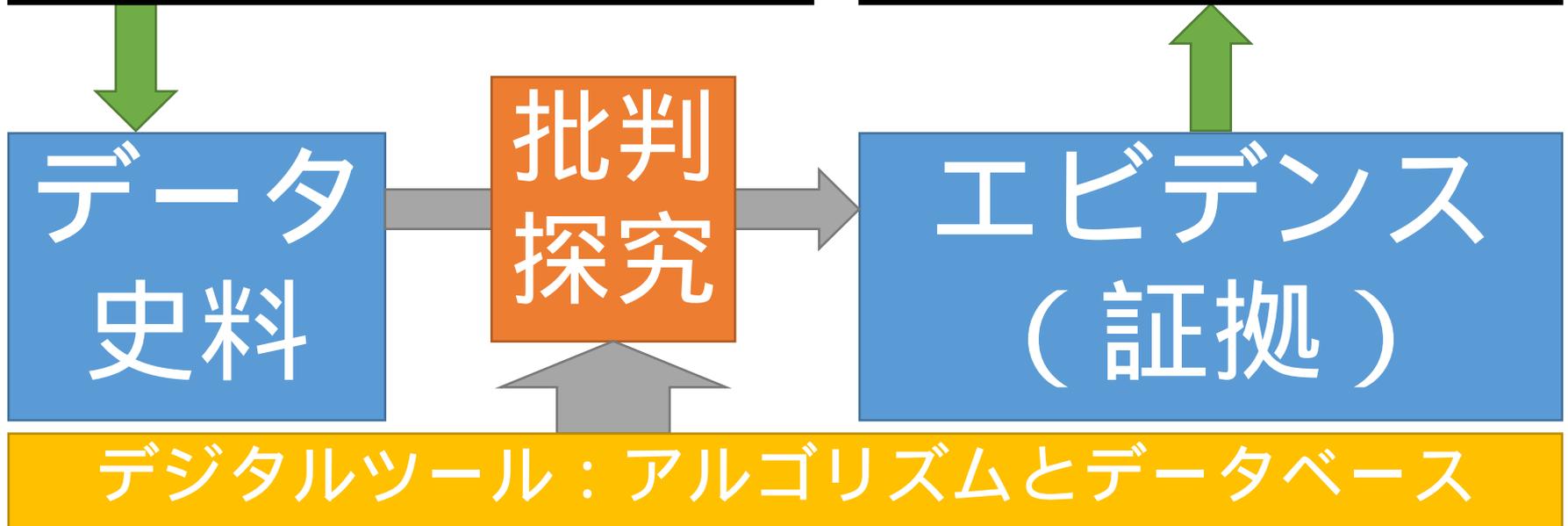
では、何をやればいい？

デジタルデータを使って
「読み方」を変える。

デジタル史料批判

数値データ
空間データ（地図）
画像データ（写真）

空間的・視覚的
証拠に基づく事
実の推定と集積

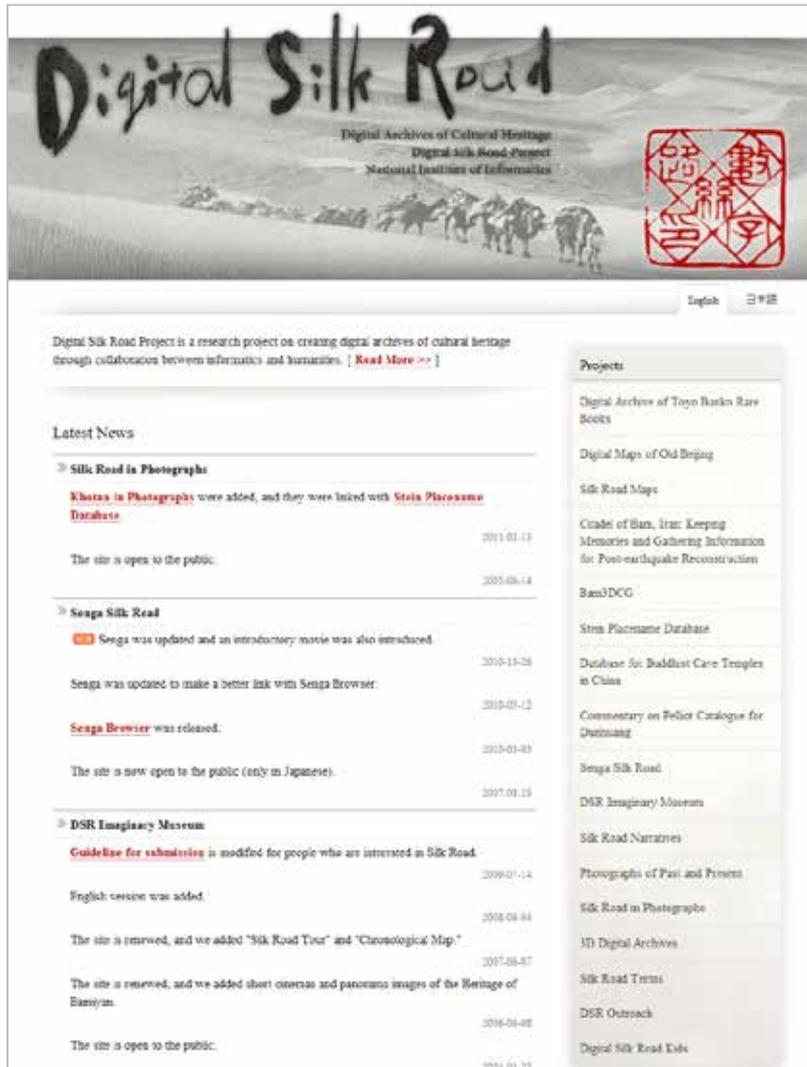


書籍デジタル化の価値は何？

- **図書館に行かなくても読める。**
- どこでも読める（フィールドワークに有用）。
- コンピュータにテキストを読ませる。
- コンピュータで定量的な分析ができる。
- データを自由に切り貼り加工できる。
- 詳細な研究成果を即時に共有・出版できる。
- みんなで協力しながら研究成果を増やせる。
- 誰もがアクセスできる研究環境を実現する。

データベースの構築

デジタル・シルクロード



- 2001年に開始。
- デジタル・ヒューマニティーズ：情報学と人文学の共同研究を通して、新手法による新知見を目指す。
- データベースやデジタル研究資源等はウェブで一般に公開。

デジタル・シルクロード・東洋文庫アーカイブ

書籍リスト 画像リスト シリーズリスト ボリュームリスト 日本語 English

ページ検索 タイトル ページ 検索 全文検索

シルクロードに関連する基本文献史料として著名な貴重書203冊（冊数76人・全59,358ページ）を、全ページデジタル化して閲覧可能なデジタルアーカイブ（画像データベース）。（[ページ一覧](#)）（[ホーム](#)）



全文検索 | 画像検索 | 地図検索 | 書籍リスト | プロジェクト | 新ことわり | 連絡先 | ニュース

Copyright (C) 2003-2013 National Institute of Informatics (国立情報学研究所) and The Toyo Bunko (東洋文庫). All Rights Reserved.
本ウェブサイトに掲載するデジタル史料の複製・転載は固くお断りいたします。

東洋文庫貴重書 アーカイブ

- 203冊の書籍59358ページを既にデジタル化。
- 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）。

画像切り出し + 画像類似検索



図の部分の自動切り出し

国立情報学研究所 - デジタル・シルクロード・プロジェクト
『東洋文庫所蔵』貴重書デジタルアーカイブ

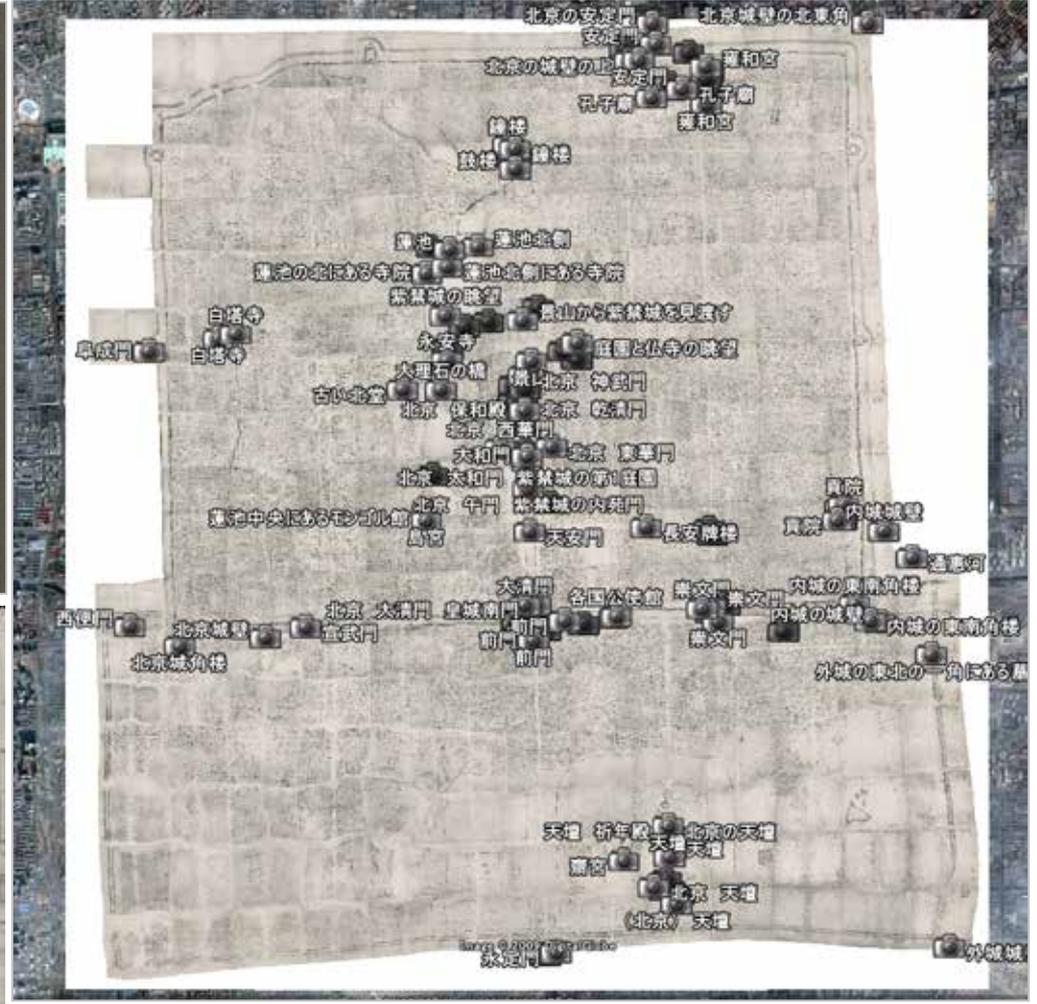
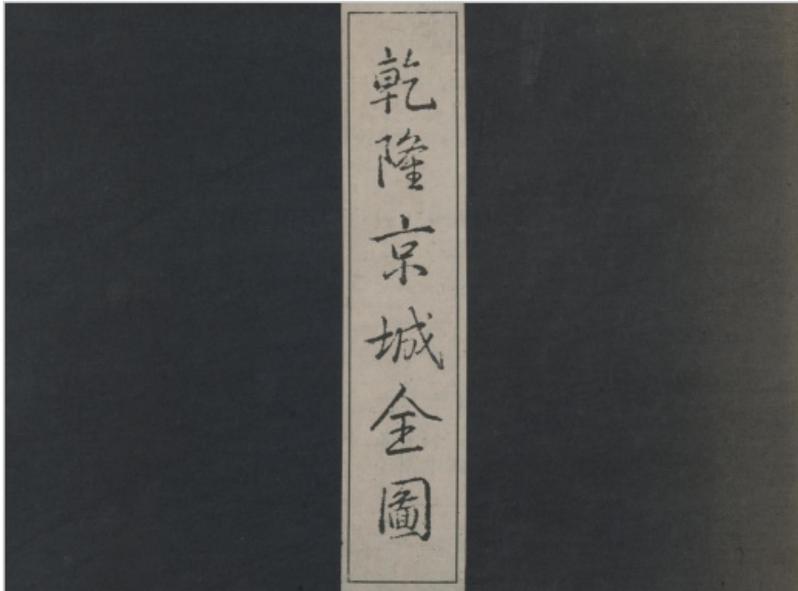
デジタル・シルクロード > 東洋文庫アーカイブ > 類似画像検索
画像検索結果 (類似検索) : 29 件 || Search in English

Query 1	1	2	3	4
VIII-5-B2-19/V-1 Page 0371	VIII-5-B2-19/V-1 Page 0377	La-56/V-1 Page 0141	T-VIII-5-A-a-3/V-3 Page 0173	VIII-5-B2-19/V-1 Page 0483
5 VIII-5-B2-19/V-1 Page 0371	6 VIII-5-B2-19/V-1 Page 0391	7 VIII-5-B2-8/V-2 Page 0287	8 Lfa-21/V-1 Page 0258	9 VIII-5-B2-19/V-1 Page 0377
10 Lfa-21/V-1 Page 0252	11 VIII-5-B2-19/V-1 Page 0385	12 VIII-5-B2-19/V-1 Page 0391	13 VIII-5-B2-19/V-1 Page 0385	14 Lfa-21/V-1 Page 0261
15 VIII-1-E-37/V-1 Page 0008	16 Lfa-21/V-1 Page 0263	17 Lfa-21/V-1 Page 0264	18 VIII-5-B2-19/V-1 Page 0403	19 XII-4-2/V-5 Page 0026

色が類似した画像を検索

空間的史料（地図）の 「読み方」

乾隆京城全圖（北京）



大規模幾何補正技術



巨大な大きさ = 縦14m×
横13m。

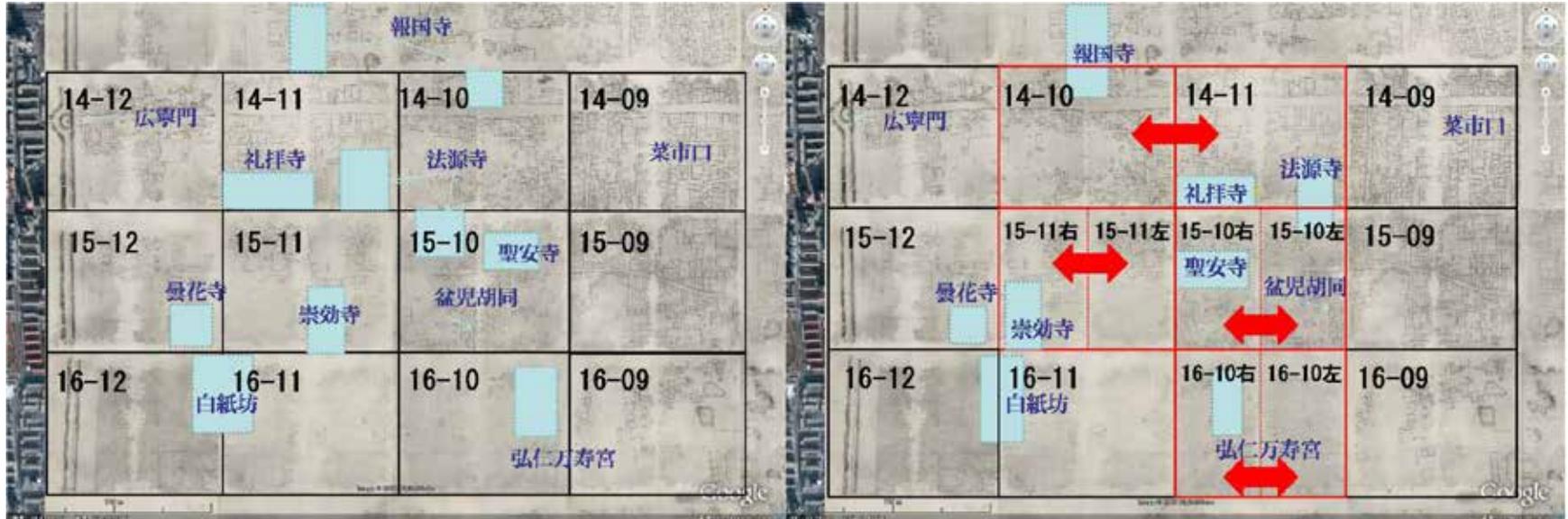
地図の分割 = 203枚の
シートを滑らかに接続。

巨大な画素数 = 全部で
290億ピクセル。

制御点 + 制御線 : 北京の
特徴である直線街路を保存する幾何補正法を提案。

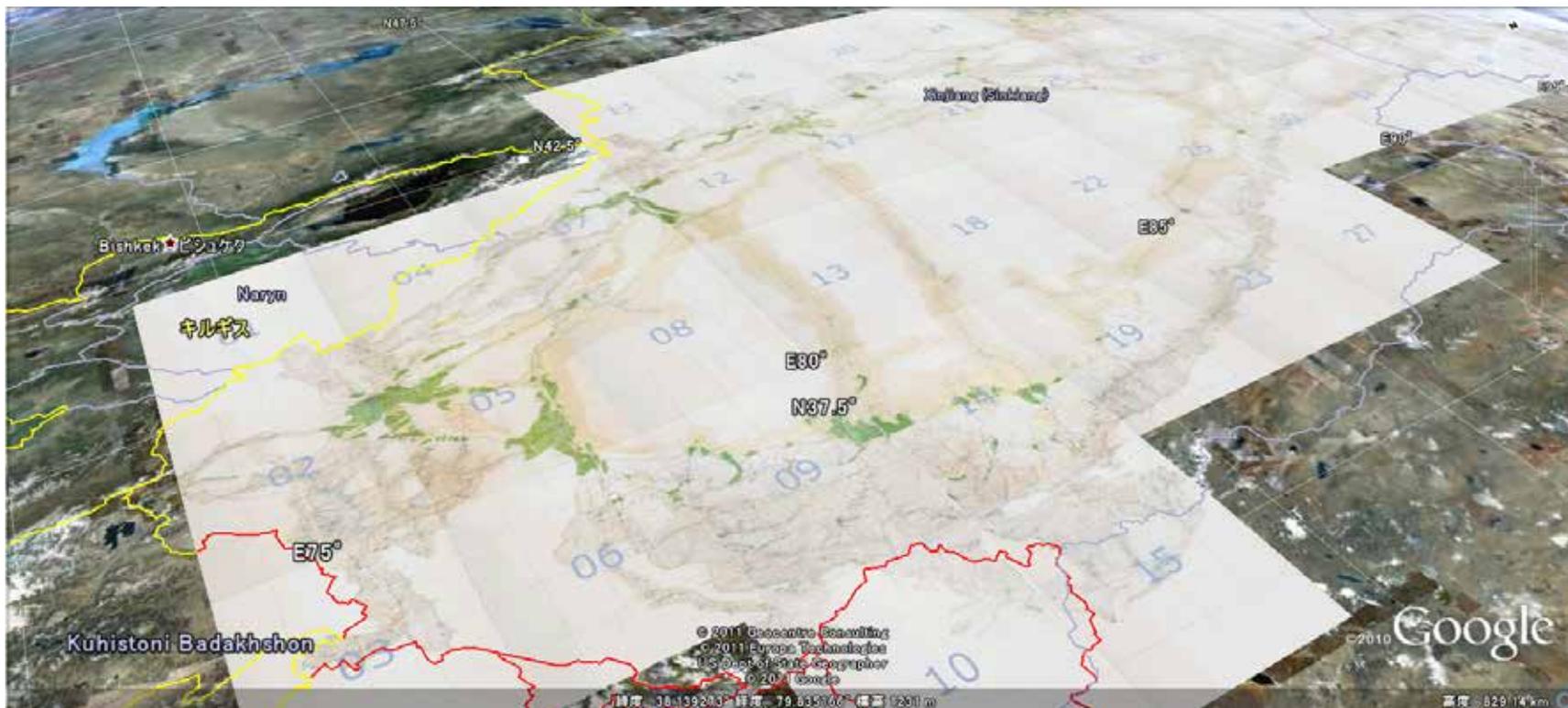


新発見とデジタル調査



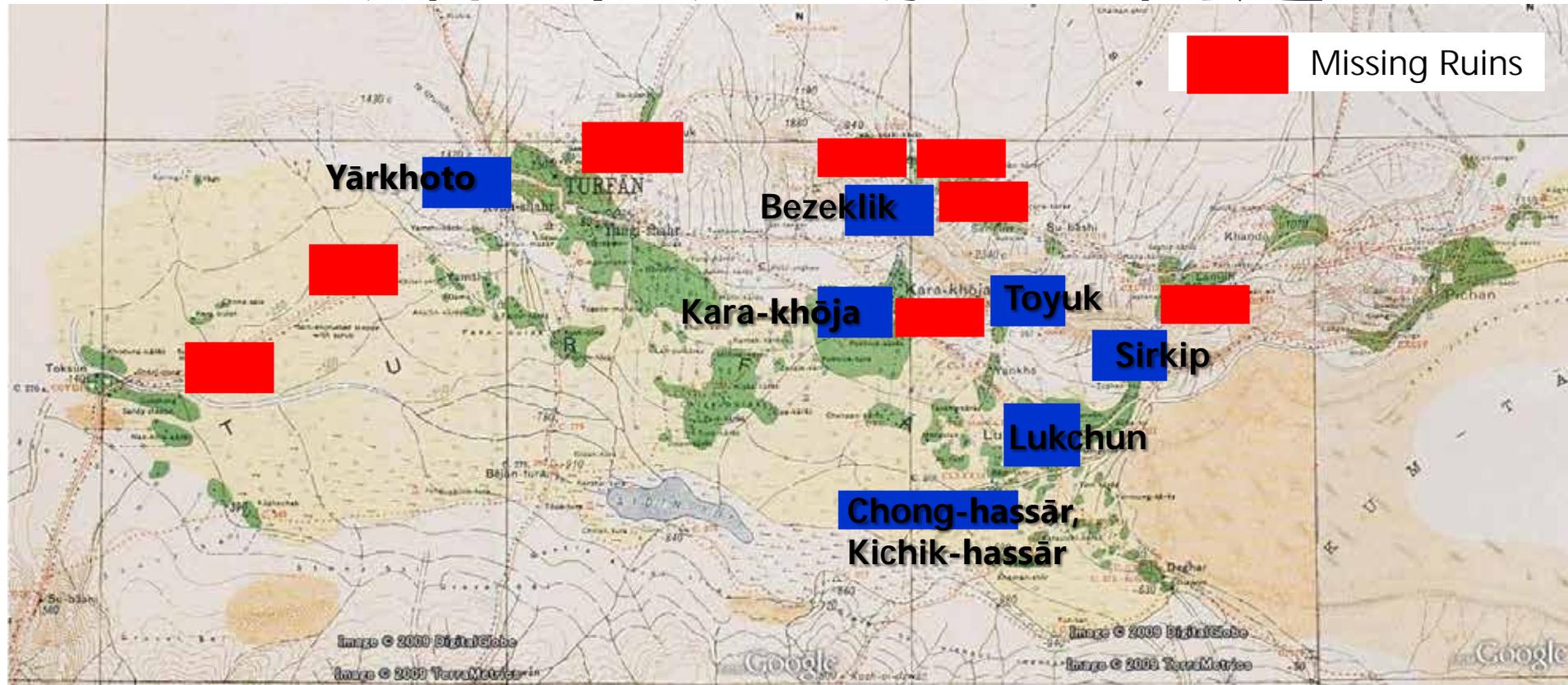
- **新発見**：5枚のシートは、過去の誤った修復（？）により、位置が入れ替わっていた。
- **デジタル調査**：地図中の地名を網羅的に収集し、従来の地名索引を増強した。

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？

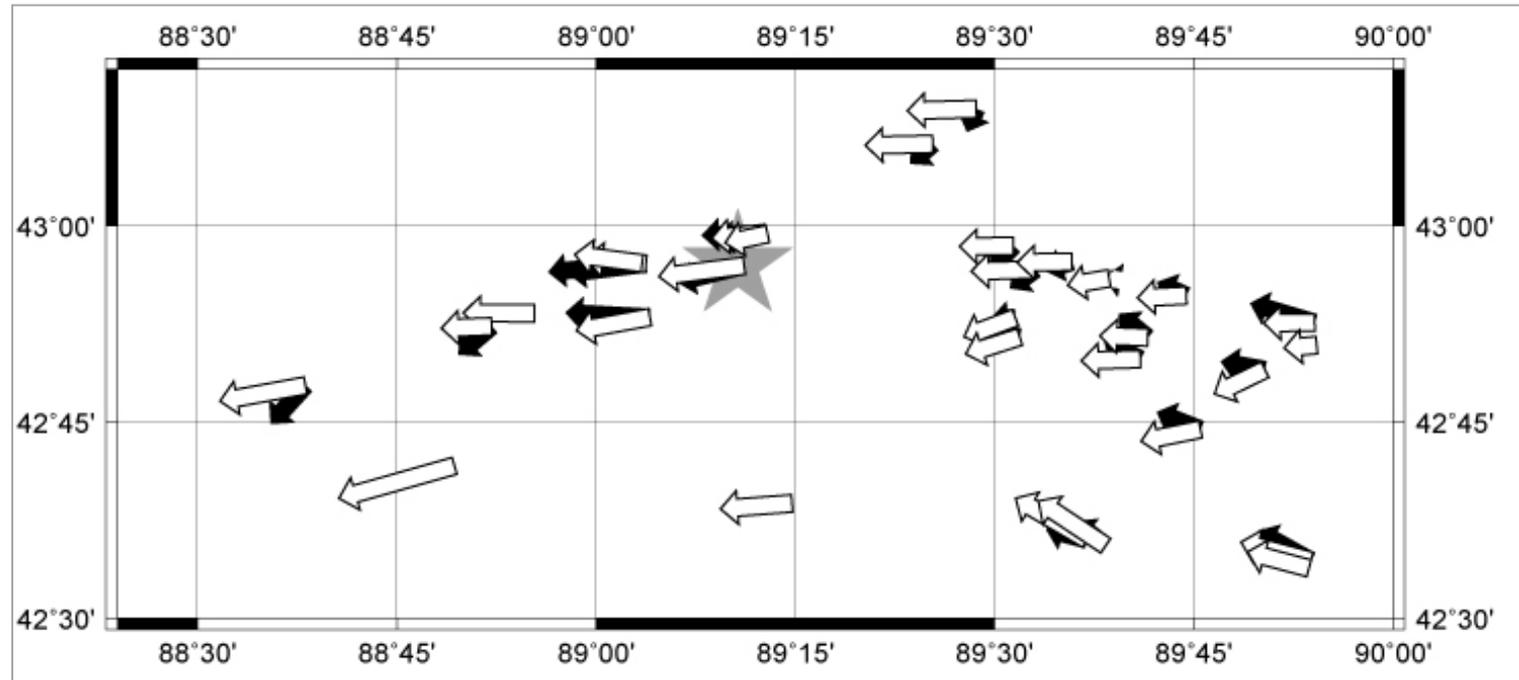
「所在不明遺跡」の問題



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

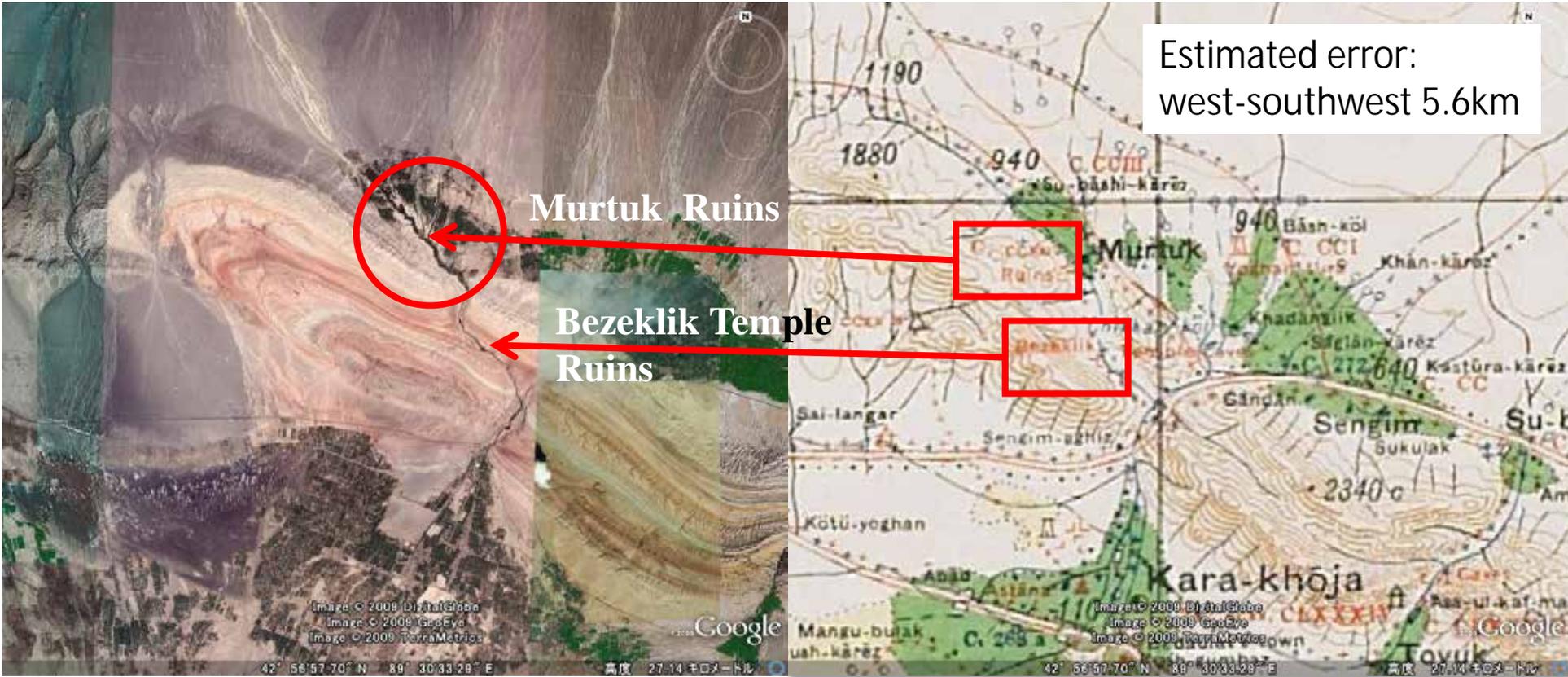
トルファンの古地図誤差分布



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。

誤差推定に基づく遺跡同定



著名遺跡の誤差推定に基づき、所在不明のムルトゥク遺跡は赤丸の位置にあると推定できる。

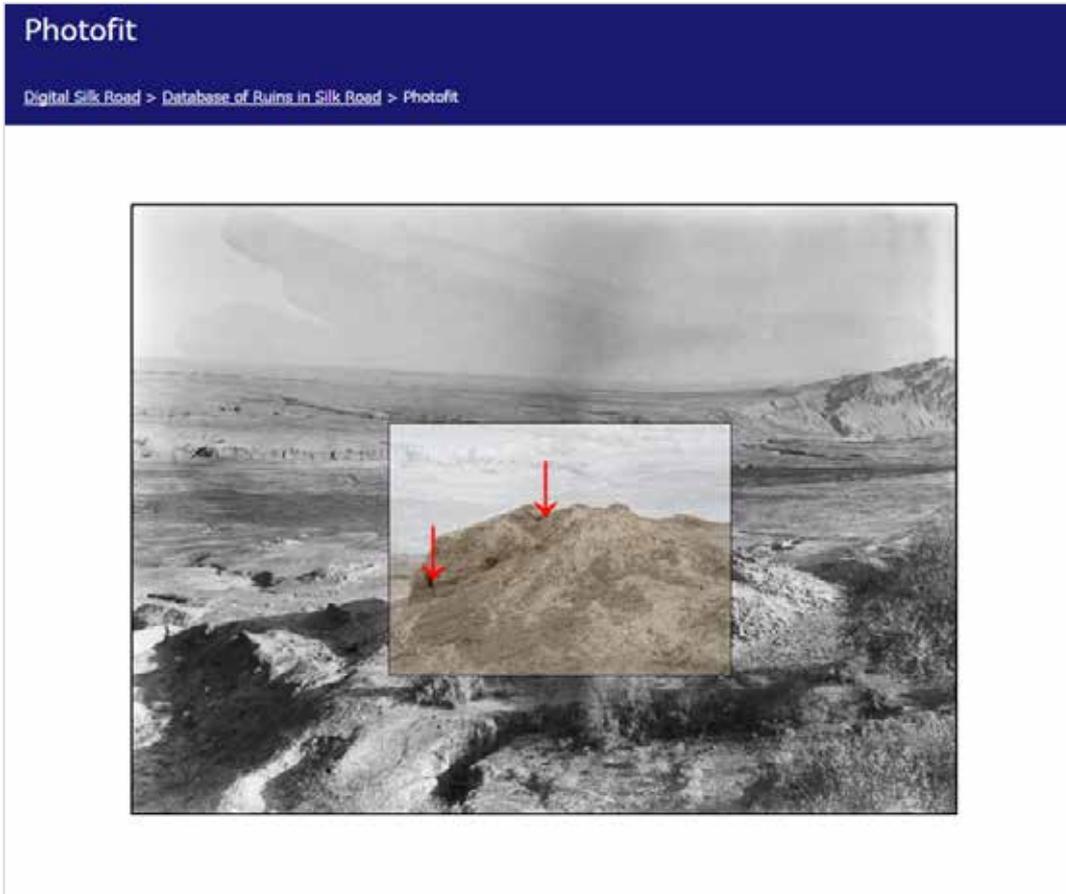
視覚的・空間的証拠 (エビデンス)の集積

Mapping : 空間的証拠



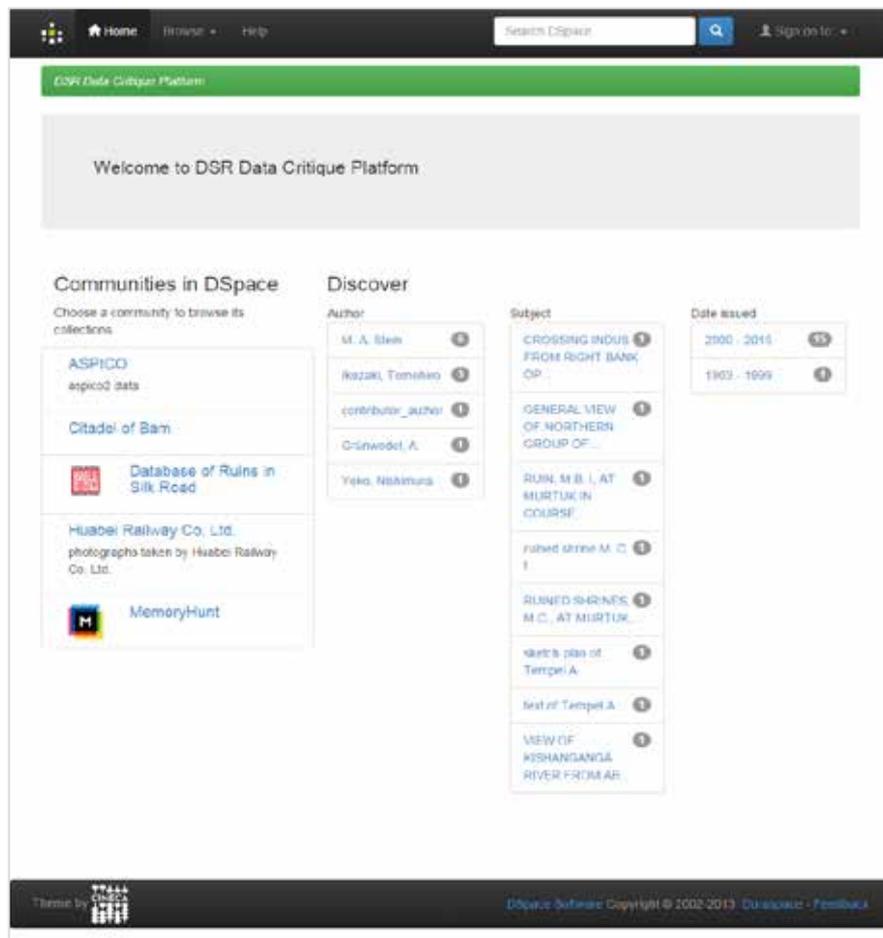
- 複数の地図の場所の一致を視覚的に確認。
- 位置あわせした結果は誰でも再現可能。
- 複数の史料を空間的に接続できる。

PhotoFit : 視覚的証拠



- 複数の写真の景観の一致を視覚的に確認。
- フィッティングした結果は誰でも再現可能。
- 複数の史料を視覚的に接続できる。

Data Critique Platform (進行中)



- 空間的証拠や視覚的証拠を集積するレポジトリを構築。
- 原著論文より細粒度・迅速な研究成果のマイクロ出版。
- Semantic Web技術を用いて、証拠の意味的な検索を実現。

フィールドワーク



メモリーハンティング

メモリーハンティング（メモハン）は、記憶の場所を探して記録するためのモバイルアプリです（現在Android版のみ）。



ビフォー・アフター写真



三宅坂より司法省を望む，写真の中の明治・大正，東京編，
国立国会図書館



古地図で司法省を探したが、解釈が複数
出て場所を特定できない。最終的には、
想定と全く異なる地点が正解とわかった。

クラウド・アノテーション



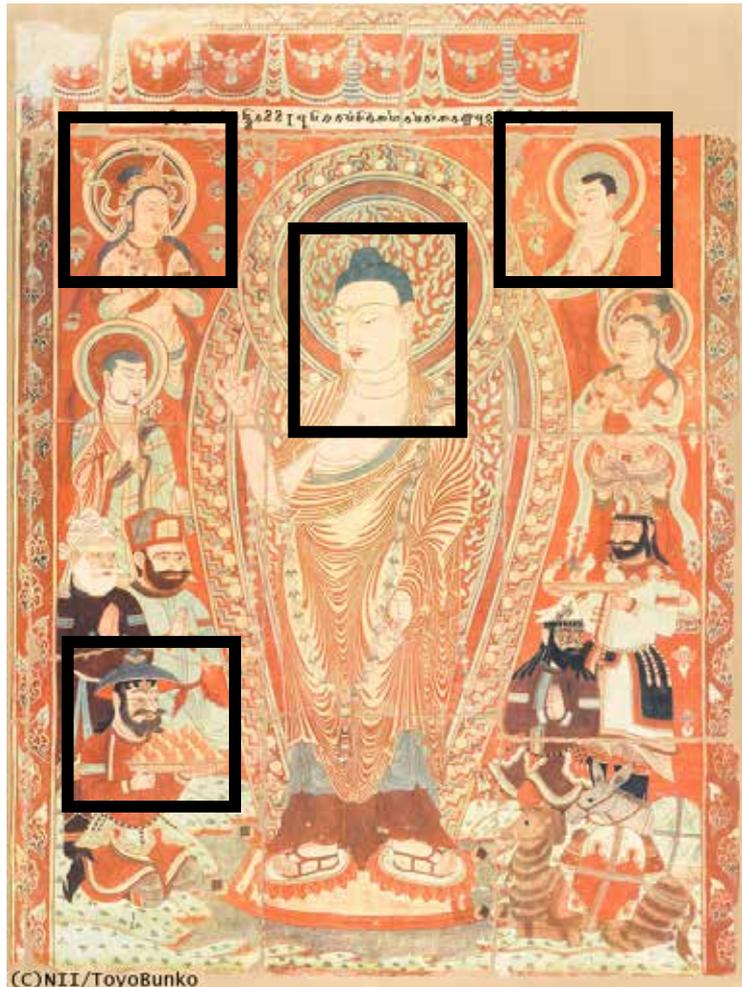
三宅坂より司法省を望む, 写真の中の明治・大正, 東京編,
国立国会図書館

研究者と市民が協力して、視
覚的・空間的証拠を収集する、
フィールドワークを実現する。



オープンサイエンス

遷画：作品の断片化と再構成



- 東洋文庫貴重書デジタルアーカイブのデジタル書籍の利活用。
- 画像の一部を断片化し、メタデータを付与して素材画像化。
- 素材画像を利用者が任意に再構成し、新たな作品として公開。

「遷画」の基本構造



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

イベントからミュージアムへ



自由に画像を並べ替え、
文化の変遷を見つけよう。

2011年10月20日、東洋文庫ミュージアムに登場。

東洋文庫ミュージアム
The East Asiatic Museum

「遼画〜シルクロード」ウェブサイト

「遼画（せんが）〜シルクロード」とは

国立情報学研究所「デジタル・シルクロード」プロジェクトは、シルクロードの文化遺産に関する資料などのデジタル化、デジタルアーカイブの構築とウェブサイト上での公開などに繋がる研究を、情報学と人文学の研究者が協力しながら進めています。

「遼画〜シルクロード」は、デジタル化したシルクロードの専門書「国立情報学研究所『東洋文庫所蔵』貴重書デジタルアーカイブ」を一般の方々にも気軽に活用してほしい、という考えのもとにデザインしたシステムです。シルクロード各地の文化遺産―遺跡や寺院的心像・壁画などを記録した数千枚の美術からお気に入りの画像を選んで並べ替え、個人の自由な発想のもとに仮想的な展示を制作することができます。

東西交易路としてのシルクロードを、文化が交差や融合を遂げつつ伝播していく様子を楽しみながら、あなたの独自のシルクロードを見つけつけてみてください。

[遼画〜シルクロードの概要](#)

東洋文庫ミュージアム

〒113-0021 東京都文京区本郷3-28-21
駒込駅（丸山手線・東京メトロ南北線2番出口）徒歩7分
千石駅（都営地下鉄三田線A4番出口）徒歩7分
上野土俗（都営/ス上58系統・英51系統）徒歩1分

東洋文庫ミュージアム

今後に向けて

画像分析の難易度を考える



- この画像とあの画像は同じものか？
- 2枚の画像は、画素のレベルでどのくらい異なるか？
- 異なり方のパターンは何通りあるか？
- この画像は何を描いたものか？
- この画像と類似した構図や印象の画像はあるか？
- ある概念はどのように描かれているか？

2方向の画像分析戦略



- **専門家の置換**：大量・高精度な作業を高速化・自動化する。
- 例：写本の同一性判定や特徴の分類。
- **専門家の補完**：少数の高度な判断を支援・効率化する。
- 例：画像のフィルタリングやタグ付け。

社会での利活用に向けて

- 社会の人々の**関心**や**問題意識**に触れれば、利活用は進む。実利が必要とは限らない。
 - @HistoricalPics : 昔の写真を投稿するだけだが、フォロワー数260万と大人気アカウントに。
- **コミュニケーター**、**キュレーター**等、関心を積極的に励起できる人が欲しい。
- (**学校・生涯**) **教育**を通して、過去から現在に至る日本文化の生の姿に触れる。
- **研究の見える化**を利活用につなげたい。

関連情報



数字
シル
クロ
ード

- デジタル・シルクロード
 - <http://dsr.nii.ac.jp/>
- 東洋文庫貴重書アーカイブ
 - <http://dsr.nii.ac.jp/toyobunko/>
- 古都北京デジタルマップ
 - <http://dsr.nii.ac.jp/beijing-maps/>
- 地図で探るシルクロード
 - <http://dsr.nii.ac.jp/geography/>
- メモリーハンティング
 - <http://dsr.nii.ac.jp/memory-hunting/>
- 遷画～シルクロード
 - <http://dsr.nii.ac.jp/senga/>
- 「デジタル・アーカイブとは何か—理論と実践」、勉誠出版、2015年6月。