

東京大学史料編纂所における データインフラ整備の現状

大向 一輝(東京大学大学院人文社会系研究科)

渡邊 要一郎, 渋谷 綾子, 平澤 加奈子

山田 太造, 山家 浩樹, 保谷 徹(東京大学史料編纂所)

目的と体制

- **東京大学史料編纂所**および**東京大学大学院人文社会系研究科**による人文学データの**共有化**・**国際化**・**連結化**
- **東京大学史料編纂所**
 - 日本史に関する史料集の編纂・出版・データベース提供
 - 編纂のための研究
- **東京大学大学院人文社会系研究科(文学部)**
 - 人文学の研究教育
 - 研究のための史資料収集
- **次世代人文学開発センター人文情報学部門**
 - 人文情報学(デジタルヒューマニティーズ)の展開
 - 知識表現・メタデータ・データベース...
 - 【発表者の立ち位置】

取り組みの概要

■共有化

- ーデジタルデータの長期保存・長期利用

■国際化

- ー海外発信の強化

■連結化

- ーデータの2次利用促進

■メタデータの検討

- ー総合データカタログのためのスキーマ整備

史料集の出版



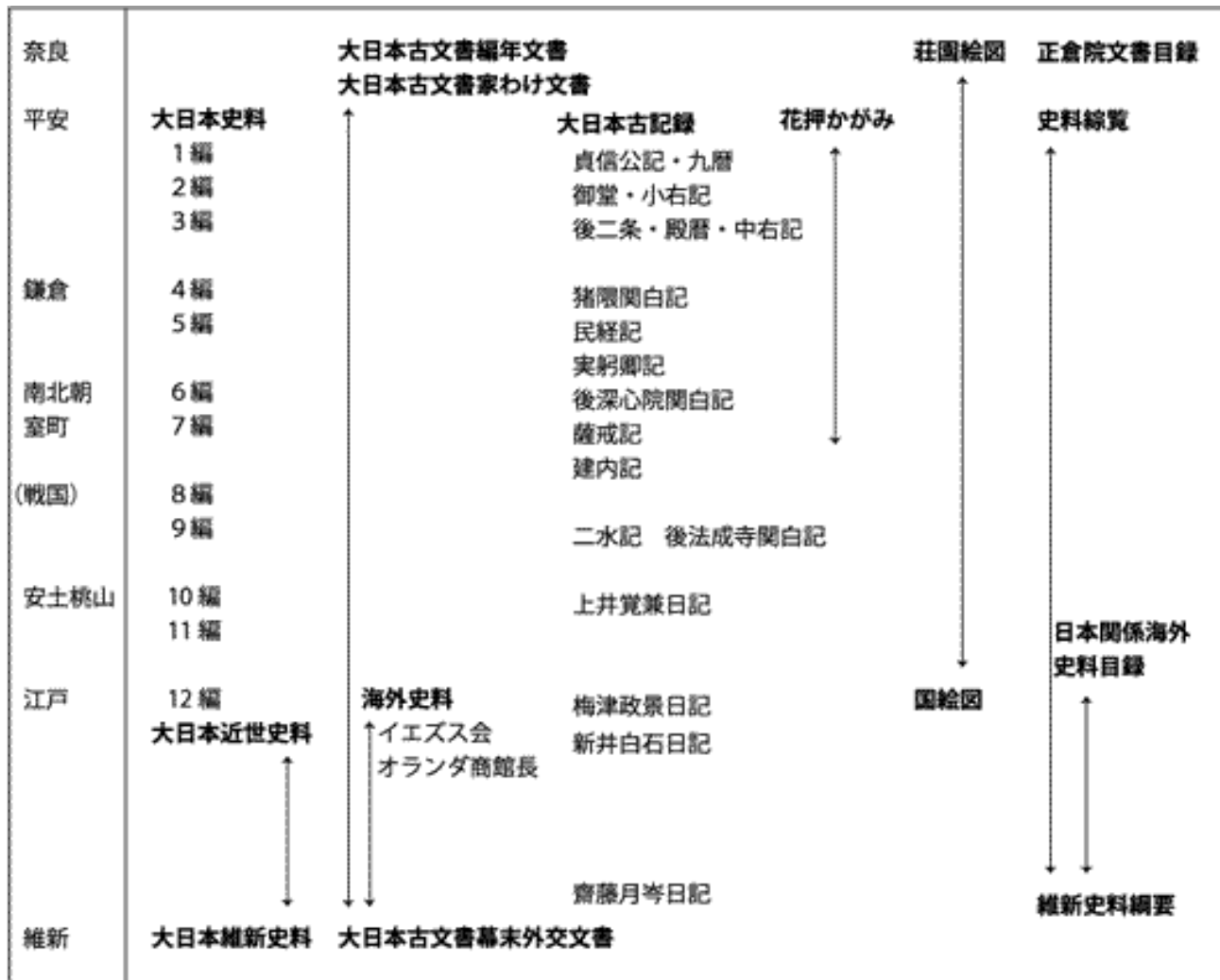
明治34年刊行の
『大日本史料』(第六編・第十二編)
『大日本古文書』(編年文書)

2つの史料集

- 編年史料
 - 事件が起きた時間の流れに沿って
関連史料を集めたもの
 - 『大日本史料』・『大日本古文書』
(編年文書)など
- 編纂史料
 - 史料群の構成と内容を精密に復元
 - 『大日本古文書』(家わけ文書)・
『大日本古記録』など

1901(明治34年)から、第二次世界大戦末期
と終戦後の一時期を除き、出版事業を継続。
年間十数冊を刊行し、総点数は1000冊以上。

編纂・出版物

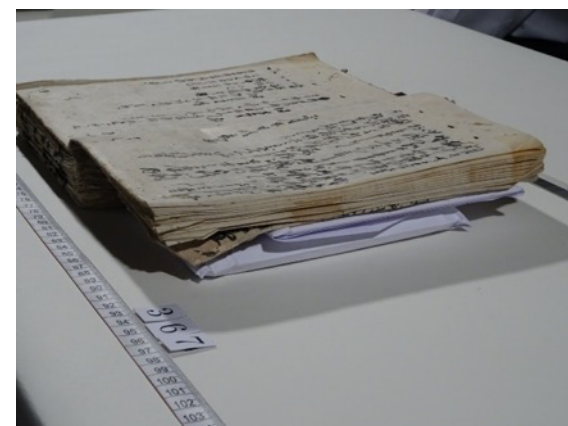
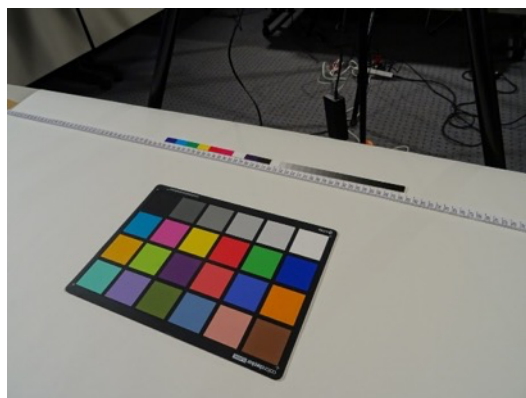


史料採訪



愛知県田原市にある田原博物館での調査.

史料採訪

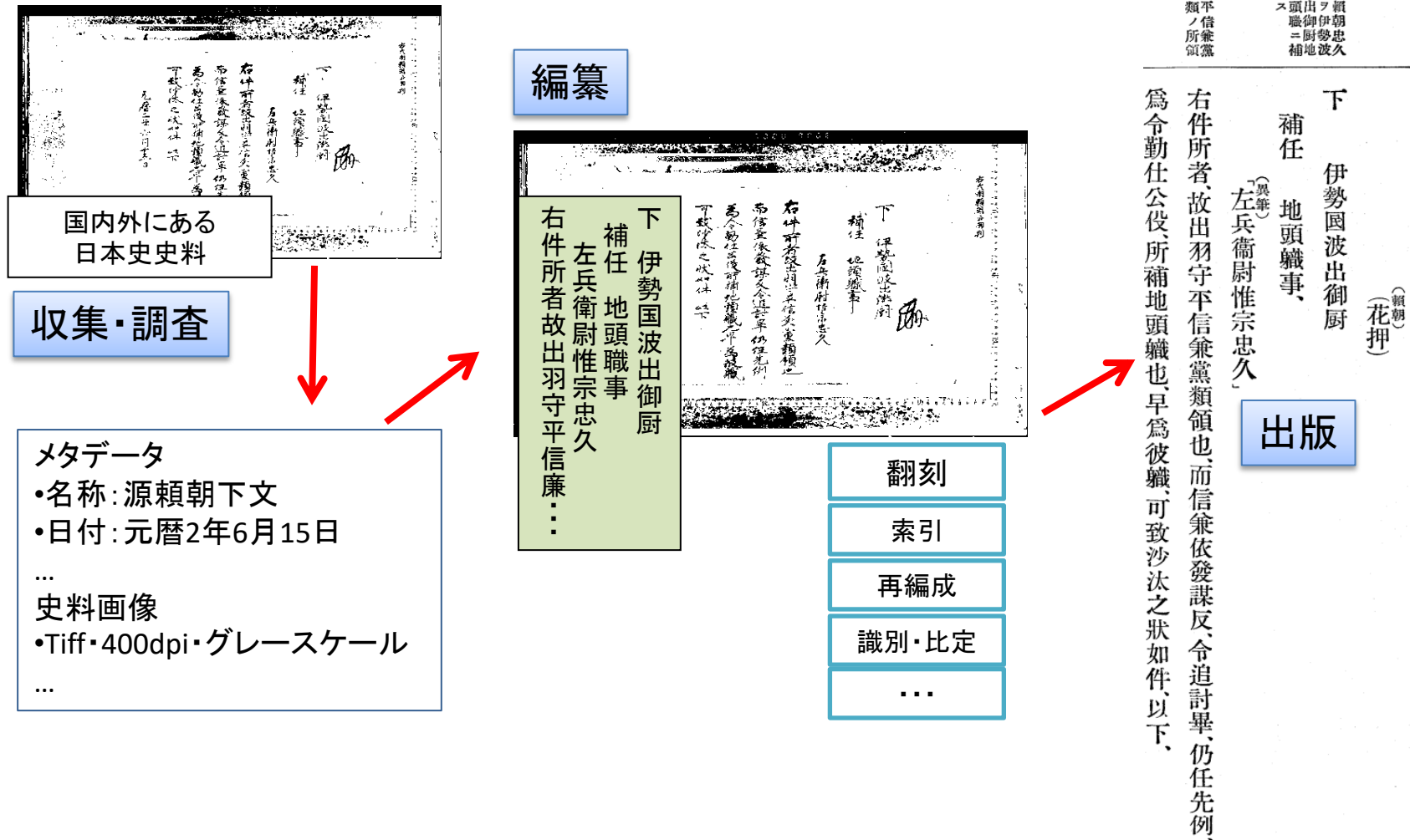


2016.11.22-24に同館所蔵「田原藩日記」(近世史料)を調査し、
嘉永～安政年間(19世紀中頃)の日記56冊を撮影した。

東京大学史料編纂所の研究フロー

対象: 前近代日本史史料(奈良～明治5年)

例: 『大日本古文書』



史料編纂所のデータベース

■史料集編纂において必要な情報をデータベース化

■種類

- 目録
- 本文
- 画像・図像・写真
- 知識
 - 人名・地名・和暦・文字...

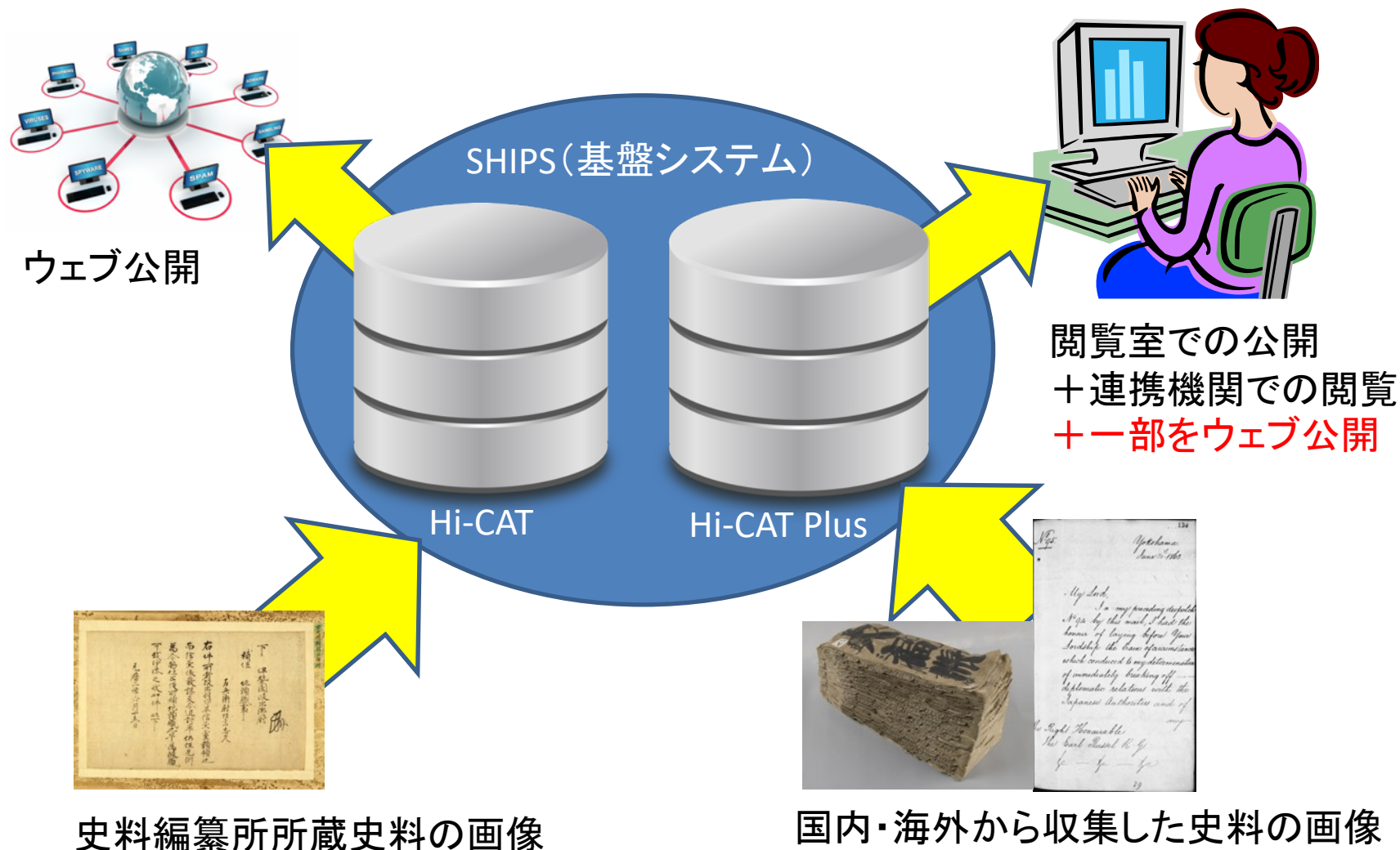
■時間範囲

- 日本前近代
 - 奈良～維新

■DB数

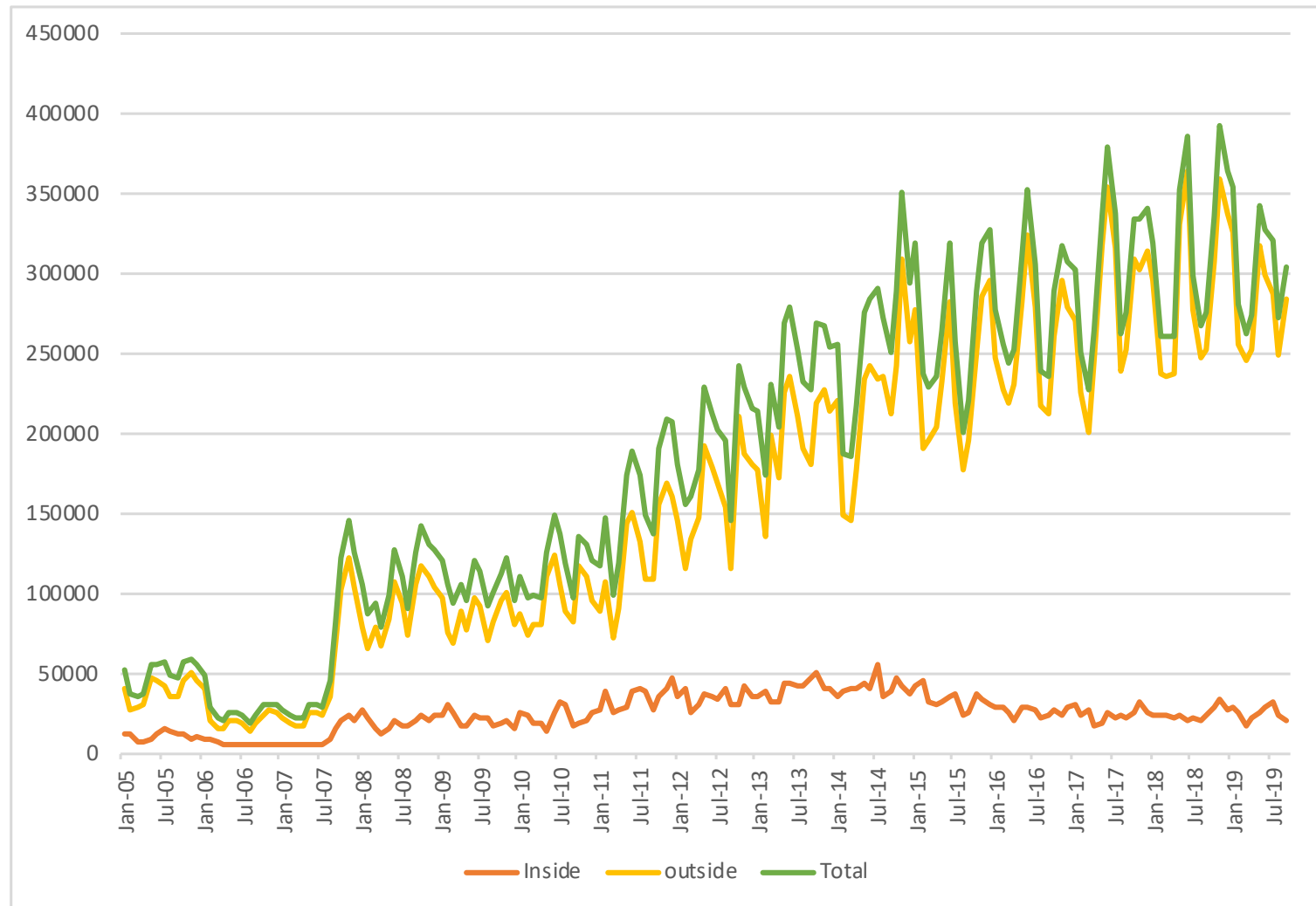
- 所外に公開: 30
- 所内のみ: 40
 - 整備中
 - 公開が困難なもの

所蔵史料と所外の史料



DBアクセス数

DB検索ページへのアクセス件数



共有化の取り組み(1/3)

■データの拡充

－史料編纂所所蔵史料

- ・備後福山藩阿部家史料: 約55,000件

－[DB選択](#)→Hi-CAT→項目検索→特殊蒐書「備後福山阿部家史料」

－所外史料

- ・天皇家・公家文庫史料(宮内庁書陵部所蔵): 約560,000件

－[ウェブサイト](#)

－[DB選択](#)→Hi-CAT Plus→キーワード「宮内庁」

－くずし字

- ・電子くずし字字典: 約283,500件

－[DB選択](#)→電子くずし字辞典データベース→詳細検索→文字「あ」

－画像からの文字の切り出し

備後福山藩阿部家史料

東京大学 THE UNIVERSITY OF TOKYO SHIPS Image Viewer

所員用ログイン

≡ 隠す

6 / 19

全画面 ホーム 拡大 縮小 左回転 右回転 ナビ表示 OFF 印刷プレビュー 一覧へ

00000001

00000002

00000003

00000004

00000005

00000006

00000007

00000008

00000009

00000010

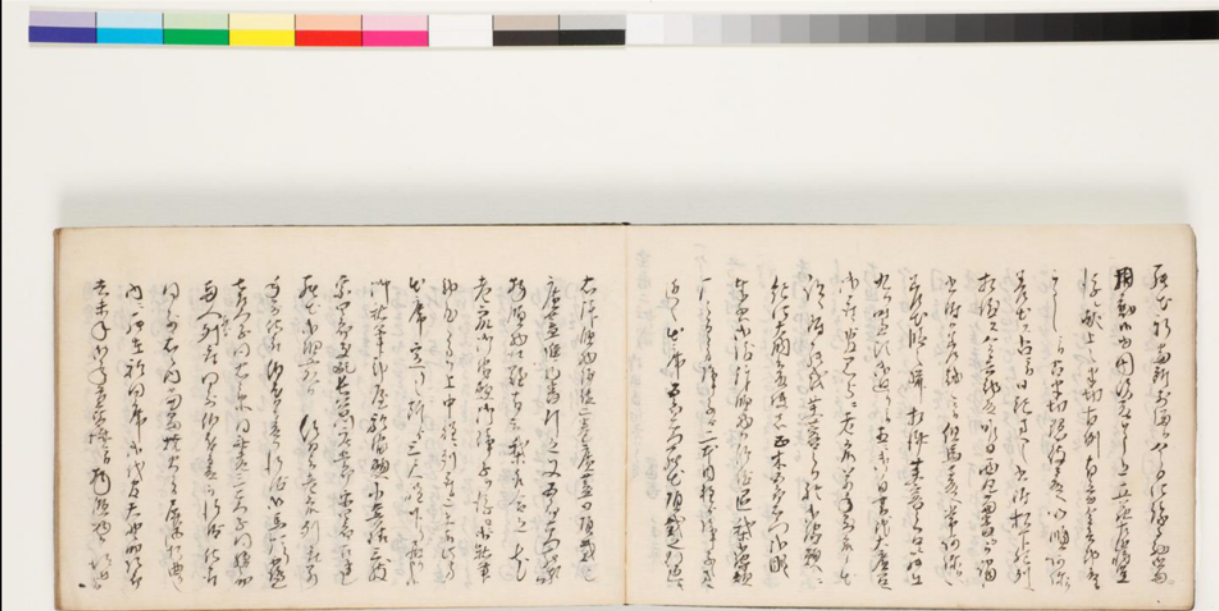
00000011

00000012

00000013

00000014

00000015



備後福山藩阿部家史料-01-001 西側入当番其外勘方留書

9 30 1 2 3 4 5 6 7 8 9 40 1 2 3 4 5 6 7 8 9 60 1 2 3 4 5 6 7 8 9 70 1

PermaLink <https://cloiimg.hi.u-tokyo.ac.jp/viewer/image/idata/T50/01/001/00000006.jpg>

電子くずし字字典

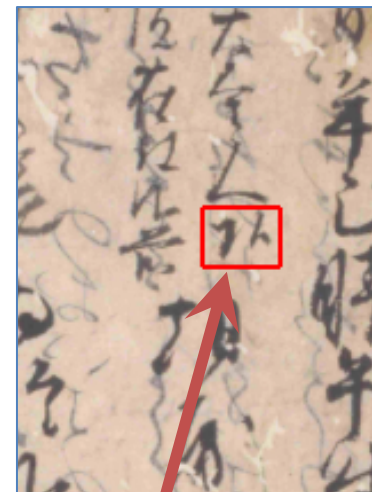
電子くずし字字典 DB選択 詳細検索 文字検索 語句検索 ヘルプ

前件 次件 先頭 最終 一覧 イメージ 類似検索1 類似検索2

類似検索3 類似検索4

1/272件

【文字】	頭
【部首】	181,151,030,025
【文字コード】	982D
【大漢和コード】	43490
【音読み】	トウ；ズ；ト
【訓読み】	あたま；かしら；こうべ
【史料区分】	原本
【架番号】	0073-10-8
【頁】	00000003
【和暦年月日】	寿永元年7月13日
【文書名】	寿永元年七月十三日条
【史料群名】	愚昧記寿永元年秋記
【差出】	三条実房
【宛所】	
【備考】	
【属性】	公家

電子くずし字字典

座標表示切替 縮尺・等倍切替 閉じる

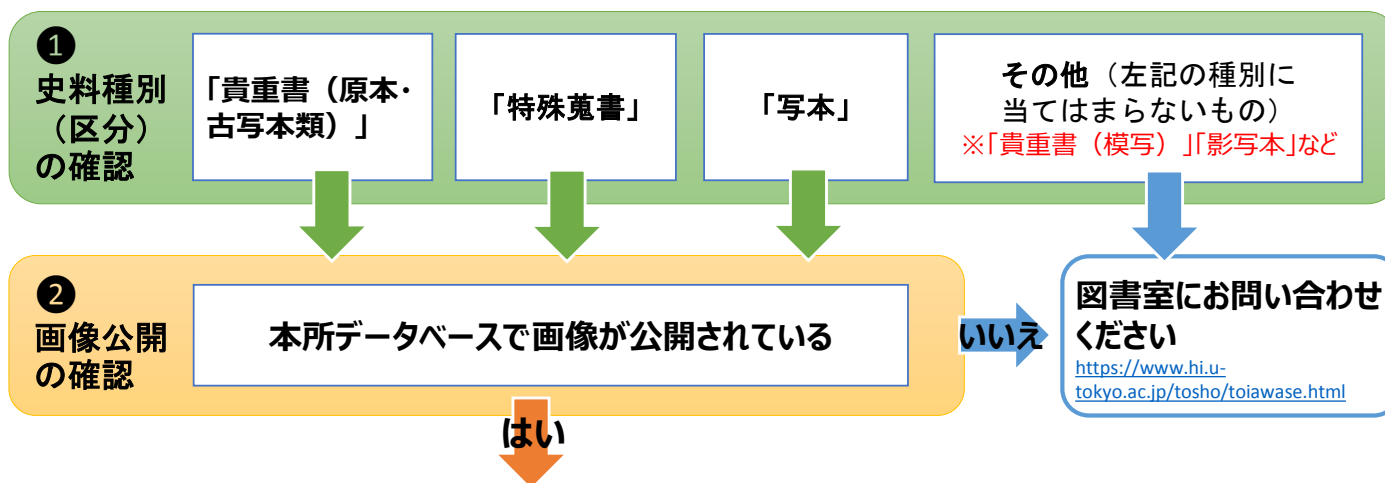
KODAK Color Control Patches Kodak KODAK Gray Scale

愚昧記 8 巻 3 表

共有化の取り組み(2/3)

■利用条件の明確化

東京大学史料編纂所所蔵史料の掲載・放映手続きについて



CC BY 4.0 相当の条件でご利用いただけます

利用の際は東京大学史料編纂所の所蔵史料であることを明示してください。

例：『史料名』（東京大学史料編纂所所蔵）

- ・ 変更を行った際はそれについて明示してください。
例：『史料名』（東京大学史料編纂所所蔵）を改変
改変した史料画像データを本所が作成・公開したと捉えられるような公開・公表は行わないでください。
- ・ 大日本維新史料稿本の場合は下記の例のような表示をお願いします。
例：東京大学史料編纂所所蔵「大日本維新史料稿本」所引「三田藩家老日記」

注

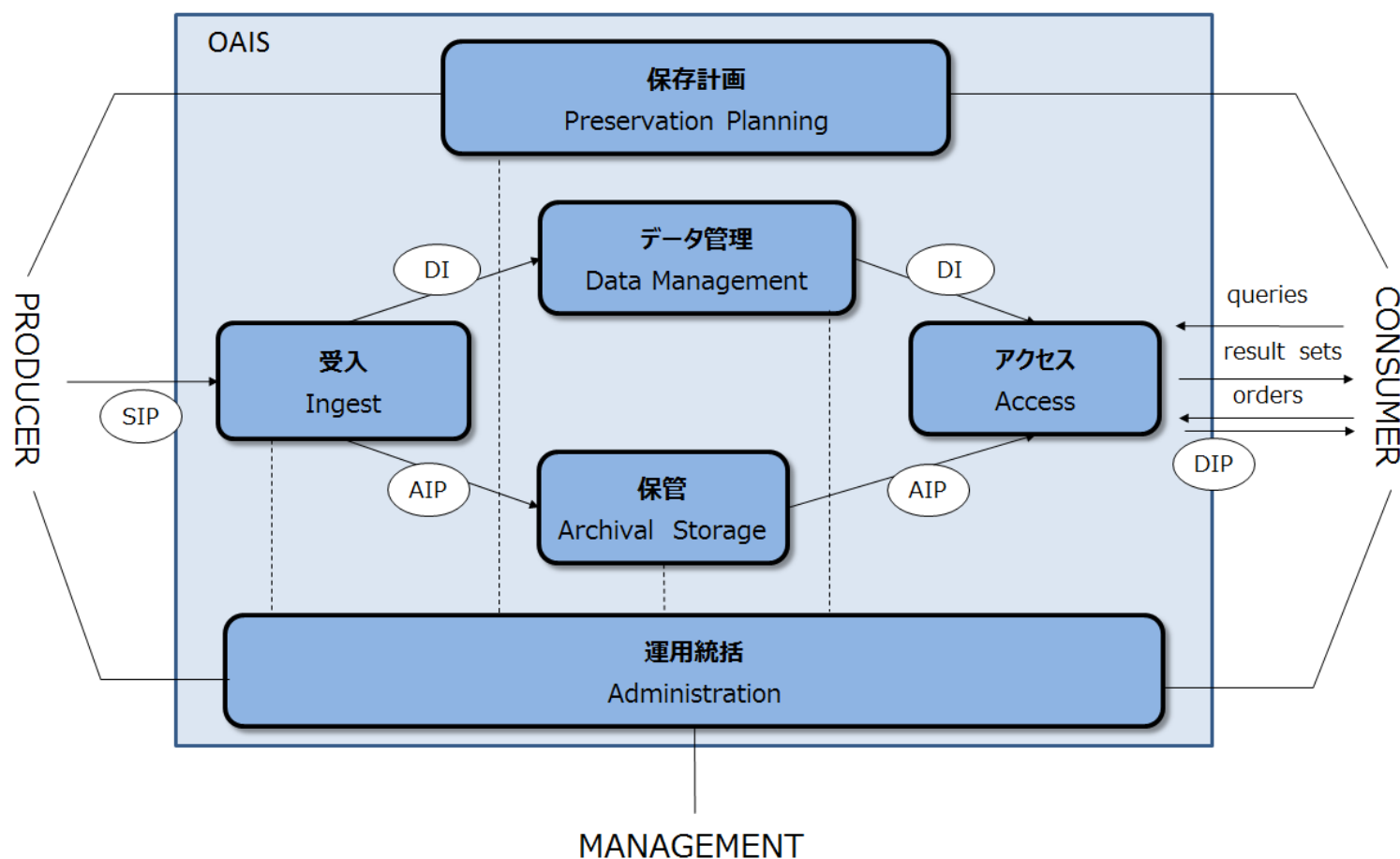
・CC BY 4.0 相当の条件に該当する史料で、本所の公開画像を使用せず、既に掲載使用されている画像（既刊書に掲載など）を使用する場合や、複製申請により入手した画像を使用する場合にも、問い合わせ手続き不要です。所蔵表示は左と同じです。

・史料種別(区分)が「本所出版物」の画像データ(「史料集版面ギャラリー」掲載)のご利用に関しては、本所ウェブサイト(<https://www.hi.u-tokyo.ac.jp/faq/reuse.html>)をご確認ください。

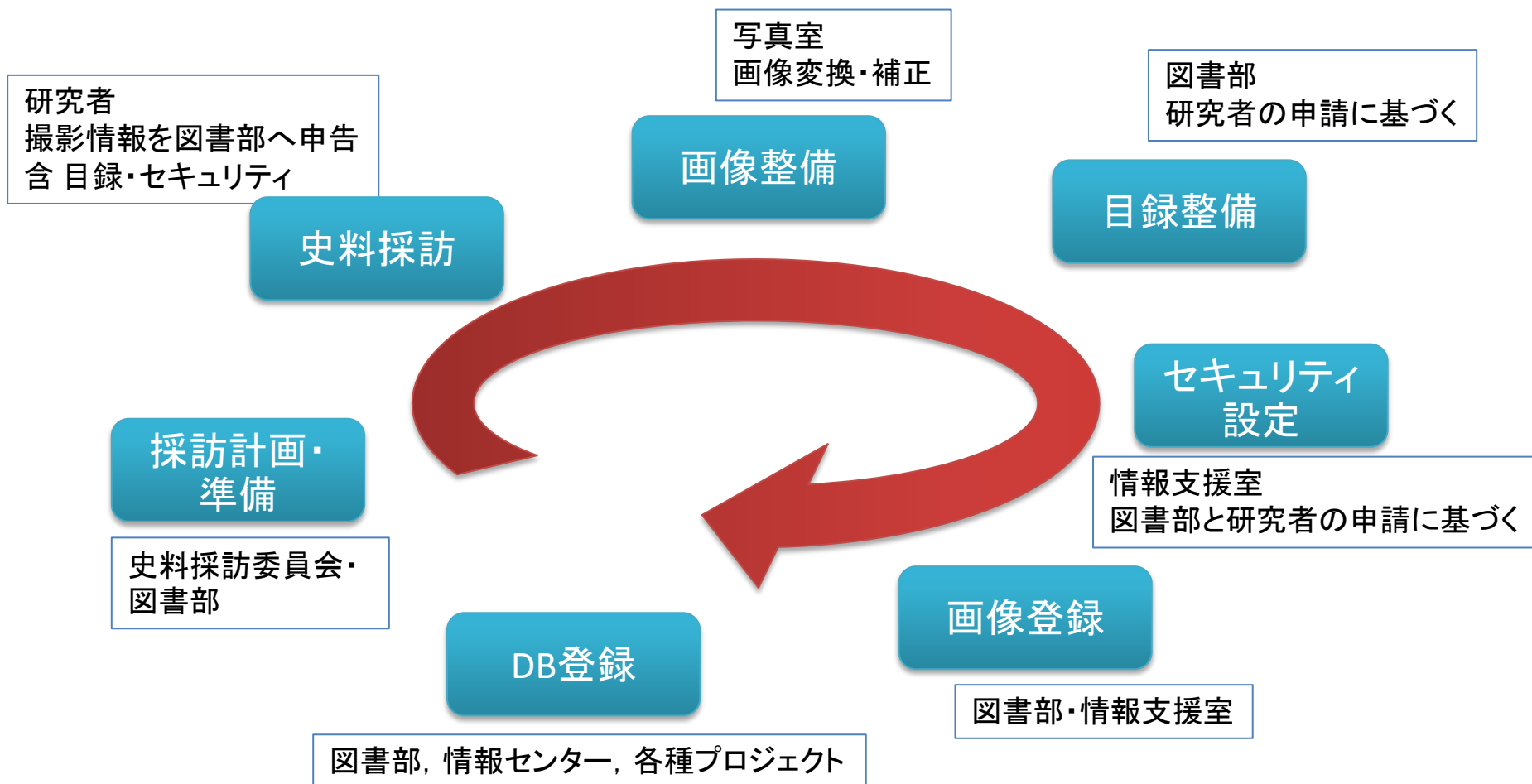
共有化の取り組み(3/3)

■ 史資料データの長期保存・長期利用システム

— Open Archival Information System (OAIS) 参照モデルに準拠



史料採訪の流れ：機関としての規約化



史料編纂所内で規約化
規約化は 2008年開始
試行は2010年, 実施は2011年から

国際化の取り組み

■データベースシステムの国際化

- 英語版ユーザインターフェースの構築

■データの国際化

- 維新史料綱要データベースの英訳
- 史料用語・歴史用語の英訳グロッサリー作成
- 日本関係史料マイクロフィルム画像の公開拡大
 - 横浜開港資料館との連携

■コミュニティの国際化

- 国際研究集会の開催(2019年12月)
- Association for Asian Studies 2020への参加(中止)

国際化：維新史料綱要DBの国際化

No	詳細	和暦年月日	綱文内容
1	詳細	安政6年1月4日	英国軍艦インフレキシブル、長崎に入港し、不正英国商人を英国政府は保護せざる旨の清国駐紮 英国公使 ボーリング及同国東印度支那日本海艦隊司令官シーモアの公翰を長崎奉行に呈す。「6日、江戸に向け退帆す。」
2	詳細	万延元年1月9日	英国公使 の通詞伝吉、其公使館前に於て刺殺せらる。
3	詳細	万延1年5月11日	幕府、英・米・仏三国公使に牒し、公使館・領事館及来航軍艦乗組員の需用を除き、本月十三日以後洋銀と一分銀との引換を停止する旨を通告す。
4	詳細	万延元年7月9日	英国公使 、将軍に謁見す。
5	詳細	万延1年10月是月	幕府、書生を 英国公使館 に託し、語学を修めしむ。
6	詳細	文久1年2月4日	仏国代理公使ドゥ・ベルクール、明日、英国特派全権公使オールコック、各幕府に牒して、公使館建設地「深川調練場」の不適當なるを指摘し、更に他所を選ばせん事を求む。
7	詳細	文久1年2月27日	英国特派全権公使オールコック、幕府に牒し、公使館建設地を御殿山「品川」に定むるに同意す。

No	Details	Date (Japanese calendar)	Kōyō Entry
1	Details	嘉永6年8月5日	Former lord of Mito domain Tokugawa Nariaki argues against the "friendly towards Russia" approach of Nishinomaru Rusui Tsutsui Masanori (Hizen no kami) and others, and urges the shogunate to reconsider.
2	Details	嘉永6年8月20日	Nagasaki Bugyō Ōsawa Teitaku reports to the shogunate that he has received official diplomatic documents from Russia and has Tetsuki Baba Gorōzaemon journey to Edo to deliver this message (he arrives at Edo on 9/15).
3	Details	嘉永6年9月20日	Sendai domain retainer Ōtsuki Heiji (Kiyotaka) submits opinions on friendly approaches to Russia to Kanjō Bugyō Kawaji Toshiakira (Saemonnojō).
4	Details	嘉永6年9月23日	On the 15th, when the formal diplomatic letter from Russia arrives at Edo, the shogunate asks the Sanbugyō to discuss the measures to be taken, and orders Jusha Koga Masaru (Kinichirō, later Chikugo no kami) to draft a reply. On this day, the shogunate orders Nagasaki Bugyō Ōsawa Teitaku to tell Putyatin that they would not be able to provide the reply in time.

連結化の取り組み(1/2)

■ International Image Interoperability Framework (IIIF) への対応

—標準規格に基づく画像配信・相互運用性の担保

- ビューアの実装コストが低減
- 異なる機能を持つ他のビューアで閲覧可能
- 同一ビューア上で他機関所蔵の画像を相互比較可能



史料編纂所ビューア



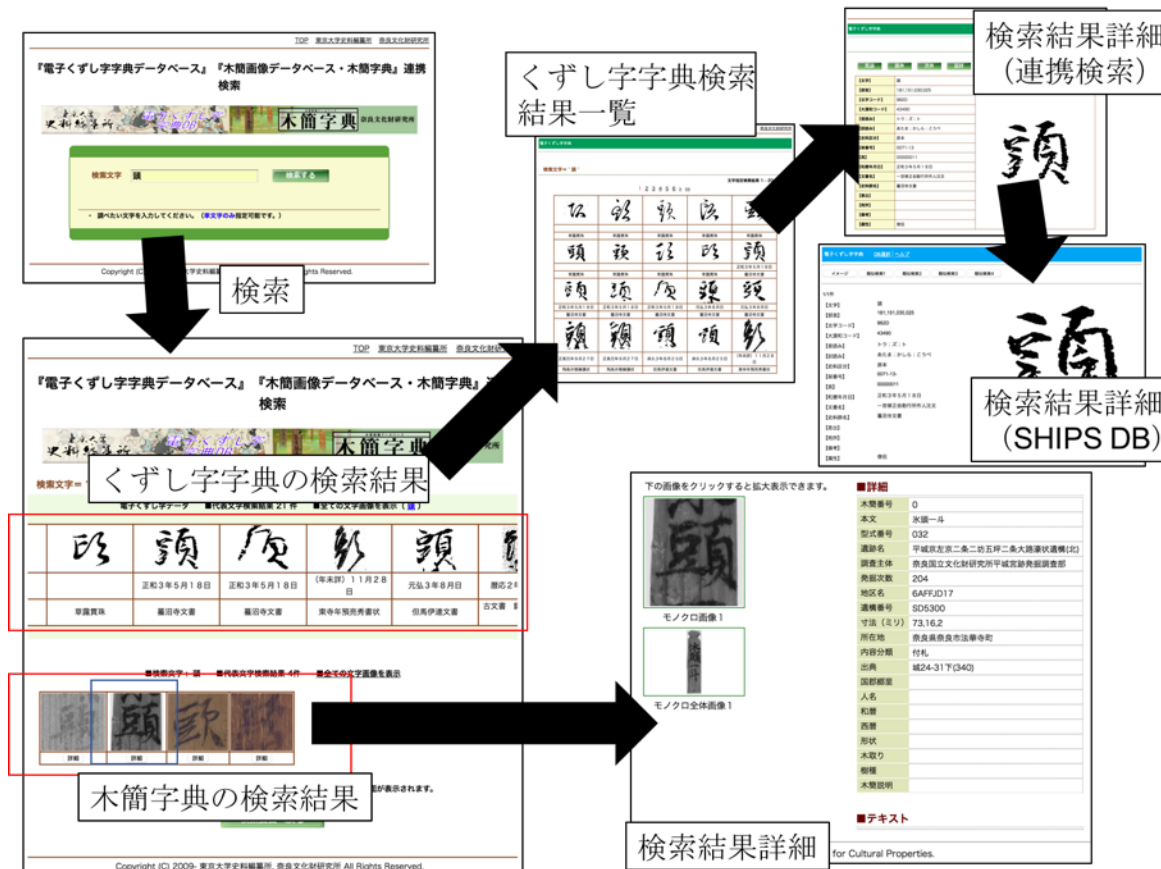
情報・システム研究機構ビューア

連結化の取り組み(2/2)

■外部システム連携の強化

—『木簡庫』『電子くずし字字典データベース』連携検索

- ・奈良文化財研究所との協働



メタデータの検討

■人文学データに対するJSPSメタデータスキーマの適用

－言語と文字

- 漢字・ひらがな・カタカナ・ローマ字

－地理的範囲・時間的範囲

- 曖昧性を有する
 - －江戸初期・寛永年間など
 - －過去の地名

－統制語彙

- データタイプ
 - －史資料・出版物
 - » 出版プロセスの介在によって区別する
 - －事項
 - » 史資料から何らかの情報を抽出し、再構成したもの
- トピック
 - －東アジアの史資料に適用可能な語彙が存在しない
 - －図書館の分類を援用

－作成者・配布者・所蔵者・寄託者

所感と課題

■ 人文学はデータが先に存在している

- 構造化手法自体が個別の研究テーマ
- 単一の枠組みに落とし込むことが困難
 - 最大公約数的アプローチ

■ データの粒度

- 基本エンティティ(情報の最小単位)は何か
 - ページ、文、単語、文字...
- 資料群と基本エンティティとの間に多様な階層性があり得る
 - 箱、冊子、巻物...
- 総合データカタログで取り扱うべきエンティティは何か
 - データベース: 数十件・画像: 数十～数百万件
 - 社会科学分野との質的・量的な相互運用性
 - ユースケースの模索

今後の取り組み

■共有化

- 史料データの拡充
- 史料管理システム上での権利関係の設定
- データ分析システムの検討

■国際化

- 史料データの英訳
- コミュニティの拡大
 - 東アジア史料研究編纂機関国際学術会議など

■連結化

- システム連携機能(API)の強化
- 総合データカタログへのメタデータの搭載