



東京大学史料編纂所における
字形データの蓄積経緯と花押
データへの展開

東京大学史料編纂所 井上 聡

はじめに①

東京大学史料編纂所とは Since 1869

「日本に関する史料及びその編纂の研究、並びに研究成果による史料集出版を行う」機関

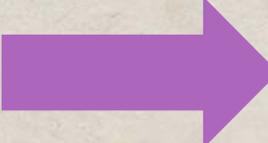
・・奈良時代～明治維新
8C 19C



はじめに②

Handwritten manuscript page with vertical columns of text. The text includes dates and lunar phase observations, such as "四月大建" (April Great Building) and "日蝕時分天陰" (At the time of the solar eclipse, the sky is cloudy). The handwriting is in cursive Japanese (sōsho).

後愚昧記 延文6年4月

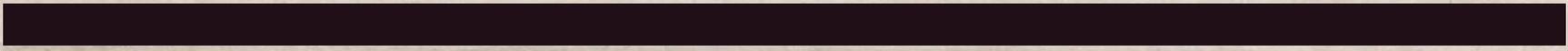


日蝕ノ曆注
 蝕否分明ニ見
 平座延引
 御願息光一人
 參仕スニ就キ
 息光ニ出フ
 公誌ニ出フ

(11張ノ紙)
 四月大建 『土府在庚』『土公在門』
 巳 月徳在庚、合在乙、月空在甲、三鏡丁美乾
 日蝕十五分之二、半、麴初中一則七十一分、加時西初刻四分、復米西六刻九分
 『天將曙』
 『水曜』 一 日辛巳 金建 延出
 大歳後、歳徳合、天恩、重

日蝕時分天陰、蝕否分明不見、平座依日蝕延引云々、參議左大弁幸圃曰人可參仕云々、其儀不審之由、

『大日本古記録』



はじめに③

基幹史料集の刊行 1901年～今日

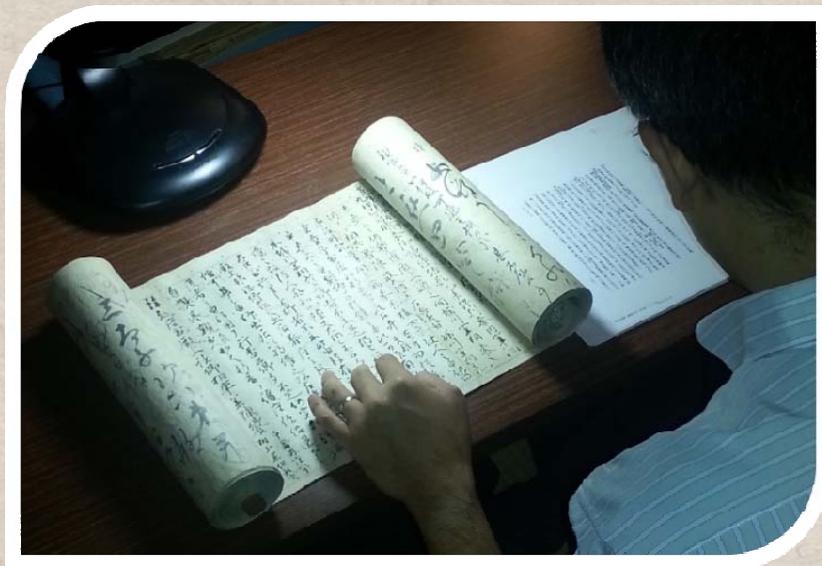
大日本史料
大日本古文書
大日本古記録
大日本近世史料
大日本維新史料
日本関係海外史料

累計 1100冊余



はじめに④

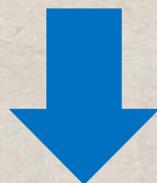
史料調査・蒐集（複製作成）
・・・1880年代～



➡ 史料画像データ蓄積 **1800万件**

I 電子くずし字字典の開発①

所内ネットワークによる史料画像の閲覧
環境が進む



難読字形を所内でデータ共有したい

2000年度 開発着手

2006年度 所内外公開へ

1 電子くずし字字典の開発②

A) 入力校正システム

目録系データベース

Hi-cat (所蔵史料)  史料画像アーカイブ

Hi-cat-plus (調査史料)

 メタデータ

 字形画像データ



登録

90°右回転

90°左回転

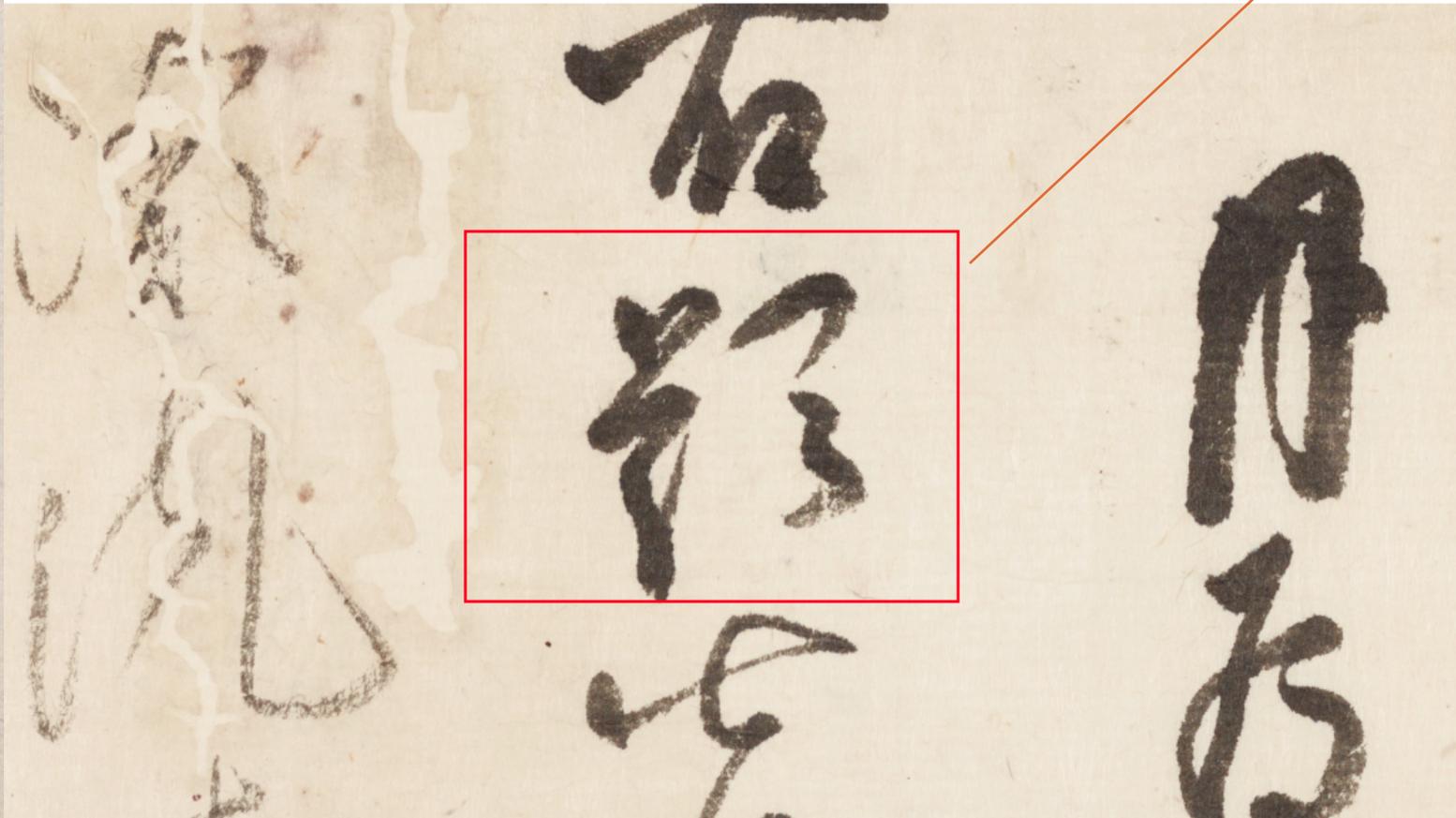
初期倍率

拡大・縮小 : 50% ▼

データベース名 : 花押 電子くずし字

ここを切り取る

画像切り取り画面



登録/更新

閉じる

管理番号 : 34288231
 和暦年月日 : 暦応元年7月2日
 西暦年月日 : 1338 年 0 07 月 0 02 日 0

和暦・西暦データ

区分 : 単漢字 ▼
 親字 : 題
 語義 :
 部首コード : 181.072.077
 音読み : ダイ
 訓読み : ひたい
 文字コード : 984C
 大漢和コード : 43584
 代表フラグ : 代表文字

文字属性データ



リポジットリ名 : idata
 史料区分 : 000
 架 : 0073
 番 : 13
 核番 :
 号 : 4
 丁頁 : 00000008
 拡張子 : jpg
 画像ナンバー : 34288231.bmp
 切取座標開始 (X) : 3697
 切取座標開始 (Y) : 1717
 切取座標終了 (X) : 4388
 切取座標終了 (Y) : 2232

元画像メタデータ

文書名 :
 巻/冊名 : 1巻
 史料群名 : 中院一品記
 差出 : 中御門宗光
 宛所 :
 備考 :
 分類 : 公家 武家 僧侶 庶民 女性 署名 在方

書誌データ

メタデータ登録画面

I 電子くずし字字典の開発③

これまでの登録データ

文字種約6000、語彙種約2600

文字画像ファイルのべ約28万件

奈良時代（8C）～江戸期（18C）におよび
133種の史料群から

Ⅱ 電子くずし字字典の機能①

簡易検索画面

電子くずし字字典

DB選択

詳細検索

文字検索

語句検索

文字推定検索

ヘルプ

文字 ▼

: 情

and or

読み ▼

:

and or

部首コード

:

and or

部首選択

■ フィールドの説明

文字 : 調べたい文字を入力してください。

読み : 調べたい文字の読みを入力してください。音読みは「カタカナ」、訓読みは「ひらがな」で入力して下さい。

部首コード : 調べたい文字の部首コードを入力するか、あるいは、部首一覧画面から選択してください（複数入力可）。

検索

クリア

Ⅱ 電子くずし字字典の開発②

検索結果表示画面(代表文字表示)

電子くずし字字典 DB選択 | 詳細検索 | 文字検索 | 語句検索 | 文字推定検索 | ヘルプ

検索結果：1件 検索式：文字='情'

No	部首	文字	画像	類似検索	連携検索
1	061,174,130,032,100	情		請清情	情

Ⅱ 電子くずし字字典の機能④

詳細検索結果表示画面

電子くずし字字典 DB選択 | 詳細検索 | 文字検索 | 語句検索 | 文字推定検索 | ヘルプ

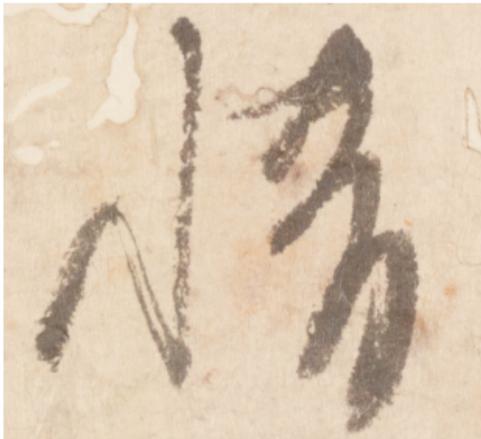
前件 次件 先頭 最終 一覧 イメージ

類似検索1

文字属性データ

1/20件

【文字】	情
【部首】	061,174,130,032,100
【文字コード】	60C5
【大漢和コード】	
【音読み】	ジョウ;セイ
【訓読み】	なさけ;こころ;おもむき
【史料区分】	原本
【架番号】	0073-13-1
【頁】	00000006
【和暦年月日】	建武5年7月7日
【文書名】	石清水八幡宮社務善法寺通清注進状
【史料群名】	中院一品記 巻1
【差出】	善法寺通清
【宛所】	中院通頭
【備考】	
【属性】	公家



書誌データ

Ⅱ 電子くずし字字典の開発⑤

さまざまな付加機能

◎類似字形テーブルの作成

電子くずし字字典 DB選択 | 詳細検索 | 文字検索 | 語句検索 | 文字推定検索 | ヘルプ

検索結果：1件 検索式：文字="情"

No	部首	文字	画像	類似検索	連携検索
1	061,174,130,032,100	情		請 清 信	信

情

請 清

No	部首	文字	画像					連携検索で検索する
1	061,174,130,032,700	情	情	情	情	情	情	情
2	085,174,074,130	清	清	清	清	清	清	清
			清	清	清	清	清	清
			清	清	清			
3	149,030,174,130,074	請	請	請	請	請	請	請
			請	請	請	請	請	請

情

清

請

類似字形表示画面

ただしアナログ作成です

Ⅱ 電子くずし字字典の開発⑥

◎「木簡庫」連携検索の活用・・・2009年10月より

電子くずし字字典 DB選択 | 詳細検索 | 文字検索 | 語句検索 | 文字推定検索 | ヘルプ

検索結果：1件 検索式：文字="情"

No	部首	文字	画像	類似検索	連携検索
1	061,174,130,032,100	情		讀清情	情

An orange arrow points from the text '2009年10月より' to the '連携検索' (Linked Search) column header in the table, which is circled in red.

『電子くずし字字典データベース』 『木簡画像データベース・木簡字典』 連携検索



検索文字 = '情'

電子くずし字データ ■ 代表文字検索結果 5 件 ■ 全ての文字画像を表示 (情)

		(年月未詳) 18日	(年未詳) 3月21日	(年月日未詳)
草露貫珠	草露貫珠	公卿門跡等書状集	得富文書	後愚昧記永徳三年紙背文書 (史料編纂所所蔵原本第十一巻)

■ 検索文字: 情 ■ 代表文字検索結果 1件 ■ 全ての文字画像を表示



詳細

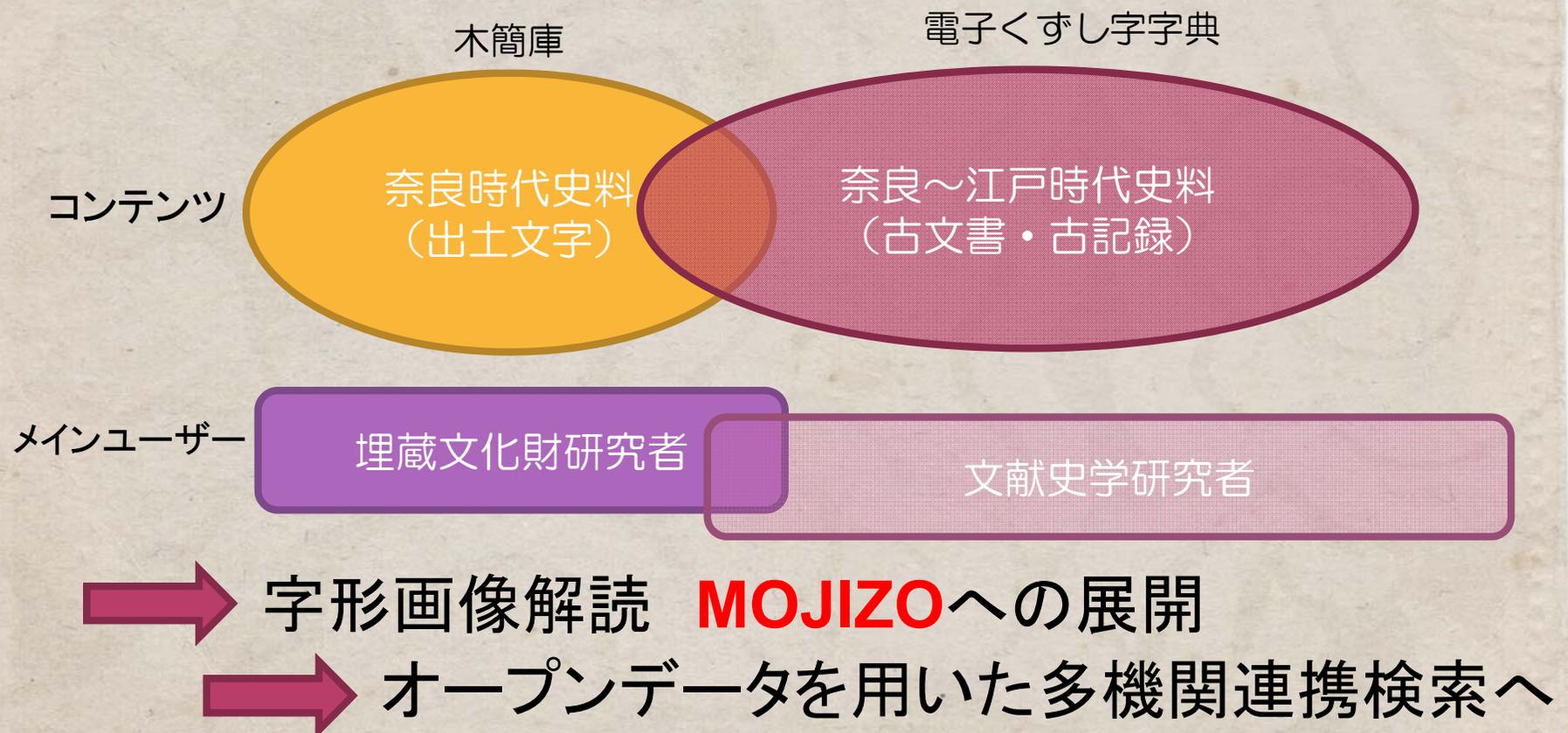
各画像下の「詳細」ボタンを押すと奈良文化財研究所「木簡字典」の詳細画面が表示されます。

検索画面へ戻る

奈良文化財研究所との連携検索画面

Ⅱ 電子くずし字字典の開発⑦

◎連携によるメリット



Ⅱ 電子くずし字字典の開発⑧

◎文字推定検索機能・・・2018年度～

史料のフルテキストデータをもとに、ある文字の前後に登場する確率の高いものを表示

電子くずし字字典 DB選択 | 詳細検索 | 文字検索 | 語句検索 | **文字推定検索** | ヘルプ

推定文字 : 前語句 後語句

■ フィールドの説明

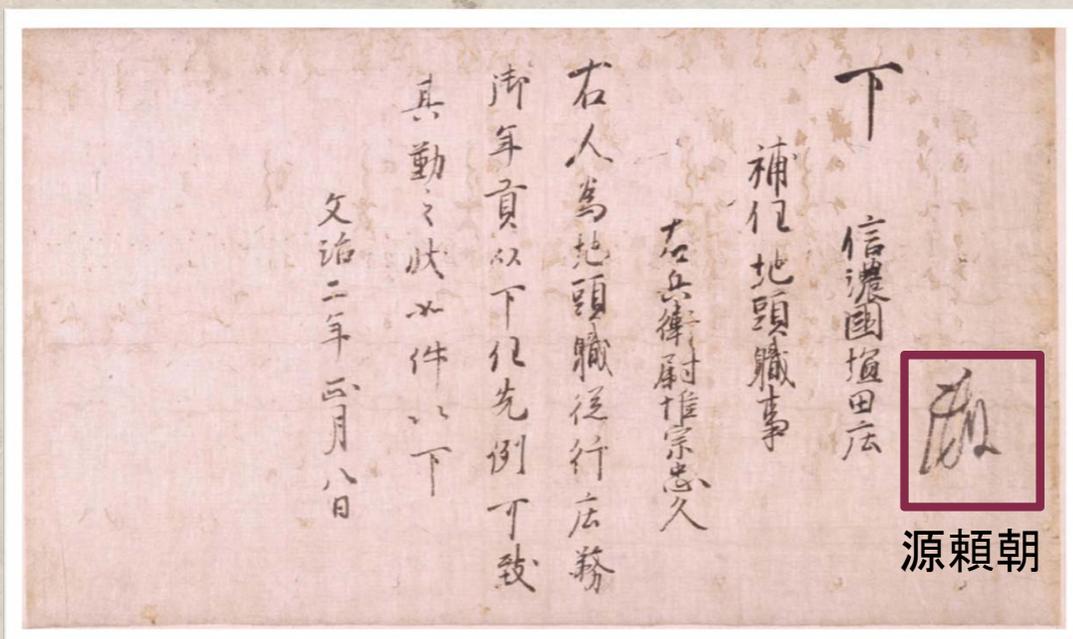
検索方法 : 調べたい推定文字の前語句と後語句を入力してください。
「 (灰色)」の部分推定文字として検索を実行します。
(例) 前語句入力: 「恐惶」
後語句入力: 「言」 → 検索結果: 「謹」

検索 クリア

Ⅲ 花押DBへの展開①

◎花押とは

文書を作成した人物が認めたサイン



オリジナルな形状
人物を特定する上で必須
文書の真偽判断にも

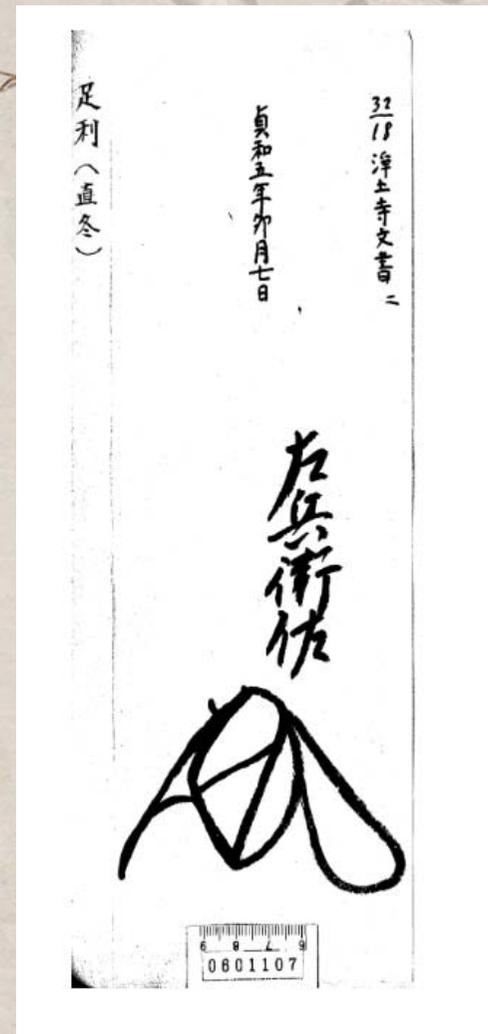
筆跡と同様に歴史的
的人物を認識する
上で重要

Ⅲ 花押DBへの展開②

◎花押収集事業

史料集編纂事業において
人物を特定する必要から、
戦前から花押を収集

➡ 「花押彙纂」



Ⅲ 花押DBへの展開③

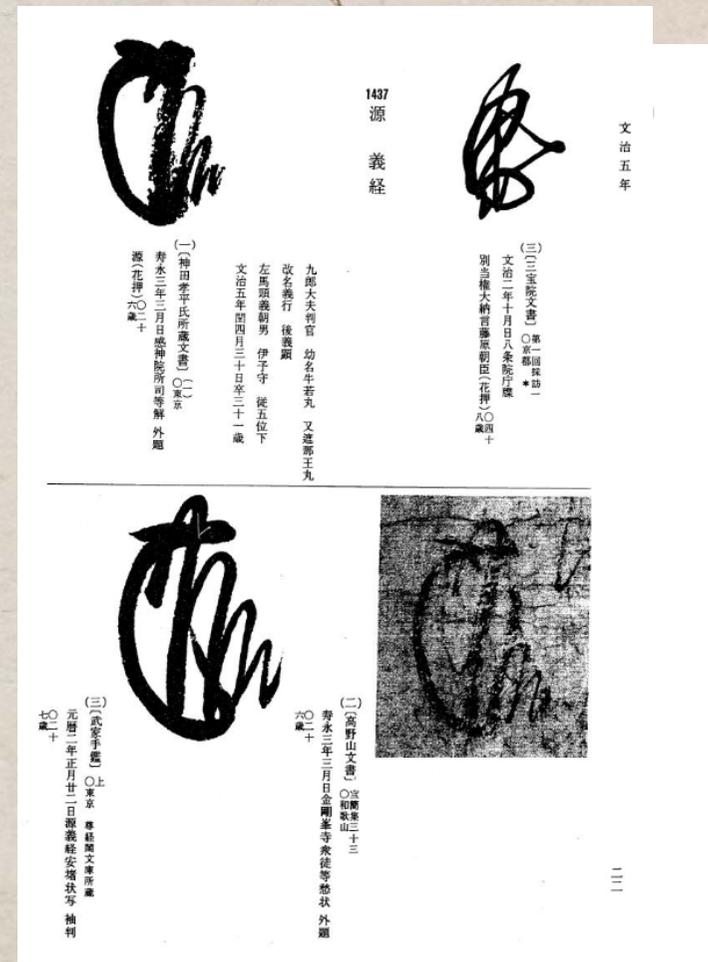
◎花押集の刊行事業

「花押カード」の作成・蓄積



『花押かがみ』

- • 1964年～，既刊8冊
- 平安時代以降に確認される人物の花押を収集し、没年に従って並べた史料集。
- • 南北朝時代までカバー



Ⅲ 花押DBへの展開④

◎花押DB構築の前提

花押彙纂DB

花押彙纂をDB化したもの・・・4万5千件

花押カードDB

『花押かがみ』編纂のため作成した花押カードをDB化したもの・・・10万件余

・・・**写し（筆写）**であることの限界

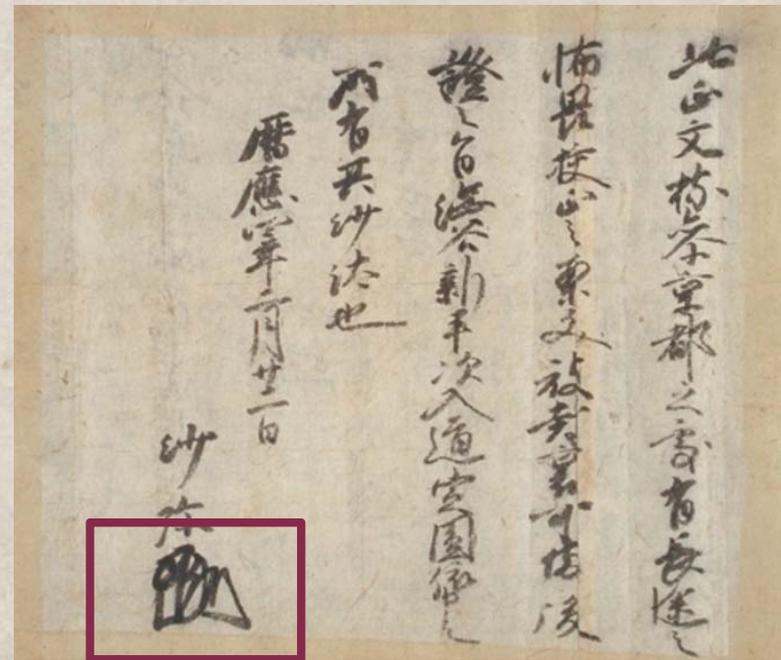
 **原本画像から花押を収集する試みへ**

Ⅲ 花押DBへの展開⑤

◎電子くずし字字典のプラットフォームを援用

①目録系DBから画像を呼び出す

②花押の範囲を指定し、花押DB
へと登録をする。



Ⅲ 花押DBへの展開⑦

◎新検索システムの構築

花押彙纂DB
花押カードDB

+

新花押データ



統合検索へ

現在、所内試験運用と
データの増強作業中

No	詳細	区分	人名	史料群名	和暦年月日	文書名	イメージ
1	詳細	カード	尊氏	北島文書	承久2年2月3日	尊氏証状	
2	詳細	カード	尊氏	北島文書	承久2年2月3日	尊氏証状	
3	詳細	彙纂	足利尊氏	上杉古文書	元徳4年2月29日	安堵状	
4	詳細	彙纂	足利尊氏	阿蘇文書	元弘3年4月29日	書状	
5	詳細	新花押	足利尊氏	歴代亀鑑(五三通)	(元弘3年)4月29日	足利高氏書状	
6	詳細	彙纂	足利尊氏	毛利文書	「元弘3」5月6日		
7	詳細	彙纂	足利尊氏	大原野神社文書	元弘3年5月16日	祈祷礼状	
8	詳細	彙纂	足利尊氏	東福寺文書	元弘3年5月18日		
9	詳細	彙纂	足利尊氏	小佐治文書	元弘3年5月18日	平基氏着到状証判	
10	詳細	新花押	足利尊氏	歴代亀鑑(五三通)	(元弘3年)6月10日	足利高氏書状	

IV 今後の課題①

1) 機械による画像学習と類似検索の高度化

* **MOJIZOの文書字形認識率**の更なる向上

桜美林大学 末代誠仁氏による取り組み

* **花押画像の解析と類似検索**の開発

埼玉工業大学 大山航氏による取り組み

➡ **オープンデータ化**による幅広い参画を促す

IV 今後の課題②

多様なメタデータを活用した発展的な検索機能

時代や地域による字形や花押の特徴・変遷

筆者の身分・性別など属性による字形・花押
の特徴・変遷

系譜や社会関係などによる字形・花押の特徴・変遷

……その他にも様々な可能性はあるはず

系譜から花押・字形の変遷を探る

父：近衛政家



連

近衛尚通



連

息：大覚寺義俊



息：近衛植家



連

息：久我晴通



さいごに

究極的に読まねばならないもの

単独の字形では読めない

記号化された語彙

文書の形式に応じた文言を予想

差出人と受取人の身分関係から文言を予測

さまざまな分野の知識を総動員することの重要性