

正倉院文書データベースにおける

検索と人文科学的研究成果との連携

後藤 真[†] 柴山 守[‡]

大阪市立大学 大学院文学研究科[†] 京都大学 東南アジア研究所[‡]

本研究は人文科学とコンピュータ研究会でも報告を続けている、正倉院文書データベース(略称：SOMODA)の開発状況を報告するものである。正倉院文書における分析の基礎的課題を抽出し、データベースによる分析によりいかなる成果があがるかを考察する。また、インターフェースの開発状況を報告し、復原のありようと、現状との比較を検討する。最後に写経所文書研究会データベースとの連携による成果との連携を示し、人文科学におけるデータベースのありようの一事例とする。

Cooperation with the search and the human sciences result of research in the ShOso-in MOnjo DAtabase (SOMODA)

Makoto Goto[†]

Osaka City University[†]

Mamoru Shibayama[‡]

Kyoto University[‡]

This research reports the development situation of the Shosoin document database (abbreviated name: SOMODA) which is continuing the report even in SIG Computers and the Humanities. The fundamental subject of the analysis in the Shosoin document is extracted, and it considers what kind of result goes up by analysis by the database. Moreover, the development situation of an interface is reported and comparison with how and the present condition is considered. Cooperation with the result by cooperation with a Sutra-Copying Bureau document study group database is shown in the last, the database in human sciences is in it, and it considers as the example like of one thing.

1. はじめに

本稿の目的は、過去にも報告を重ねてきた、正倉院文書データベース(SOMODA)の開発状況を報告するものである。正倉院文書データベースに関しては、本研究会においても数度の報告を重ね、議論を積み重ねてきた。正倉院文書はその複雑な文書の形成過程により、研究に際して、分析・注意すべき点が多岐にわたっている。同時に複雑かつ多様な資料を参照せねばならず、日本古代史研究において、正倉院文書を使う際、敷居を高くしているといえる[1][2][3]。この複雑な資料の統合というメリットに関しては、過去に報告している[4]。同時に、日本古代史料として、第一級のものである正倉院文書は、古代史科学の確立にむけて、もっとも重要な材料であることは言を俟たない[5]。そのため、XMLによりデジタル化することによって、その特徴を導き出し、特殊な部分を見つけ出すことによるあらたな学問的知見の可能性についても過去に言及してきた[6][7]。

また、あらたな研究の可能性として、古代文書の形式に即したデータ入力の可能性も言及した[8]。

本報告においては、上記の成果をうけて、これらの実現のための具体的なシステムの開発状況とその課題について報告したい。

2. 情報の統合と検索システム (1)

先述したとおり、正倉院文書においては、同時にいくつもの情報を参照し、確認する必要がある。以下、要素ごとに列挙し、課題をあげる。その後、SOMODA における表示のしかたとその分析の方法を紹介したい。

(1) テキスト本体、およびそれに付随する情報。おもに文献歴史学が主要な対象とするテキスト本体の情報である。

(2) 画像情報。図像による詳細な文書の情報は、現在の歴史学の水準であれば必須の情報であるといえる。正倉院文書の多くは造東大寺司写経所に関連する文書が多い。その文書がどのような筆致で書かれているか、精密な筆致なのか、殴り書きなのか、文書のどの部分に記載されているのかなども重要な情報である。また、写経所などでは、文書などをやり取りした、包み紙や、経典作成の際に整形した紙の一部をも反故紙として用いることがある。これらの情報は、画像でしか確認することができない。また、明治20年代～30年代において、正倉院文書統々修の編成を行ったときに、付されたとされる「付箋」の情報も現在は、画像にしかない情報である。これらは、日本古代における他の史料と決定的に異なり、史料分析の前提として画像情報が必要であるということを示す。

(3) 目録情報。接続の煩雑さについては、以前より指摘していることである。それ以外にも、複雑なのは、文書の使用状況も指摘できる。正倉院文書は背面を用いることによって、その複雑な現状があることは、接続の関係で指摘してきた。そのことのみならず、文書の二次利用は、正倉院文書の来歴と経緯を分析するとともに。当時の帳簿の構造なども分析することにつながるのである。たとえば、ある帳簿は、天平十年代の民部省関係の文書を再利用して帳簿化し、さらに、“アーカイブ”用の帳簿として作成している。このような文書では、二次面と一次面の関係は大変煩雑であり、これらを分析することは、当時の帳簿作成のありかたを詳細に示すこととなる。

(4) 料紙情報。料紙に関する情報は大きさ・料紙の質・界線が引かれたかなどの状況が主要な課題となる。一般に古代の紙は幅50センチ前後が規格であり、この大きさではない紙は、何らかの特殊事情があることになる。また、今までに詳細な研究はないものの、どのような帳簿に界線が引かれていたのか、その引き方はどのようなものであったのかなどといった文書の機能と形態に関わる点なども今後分析対象となってくるであろう。

(5) テキストの生成のありかた。現在、正倉院文書に関して、用いる基本的なテキストは、『大日本古文書』（編年）25巻である（以下は『大日古』と略す）。しかし、この『大日古』には原テキストから翻刻するとき、読み間違いが多くあることがすでに指摘されている。特に1～6巻は、正倉院文書の写本を用いて翻刻されたため、間違いが多い。一つは、写本作成の際の読み間違い、もう一つは写本より転写・翻刻の際のミスである。正倉院文書は近世末～明治期までの「整理」の際、写本も多数作成されている。また、『大日古』は初版が明治30年代であり、それ以来、長きに渡り版を重ねている。しかし、「版によって史料の文字が異なるのは、引用されるべき史料として望ましくない」という理由で、いかなるミスが翻刻に存在しようとも、変更を加えないという方針が貫かれている。つまり、『大日古』は明治30年代の状況をそのまま反映し続けているのである。『大日古』は当時の理解の水準において、若干文書の復原を試みている。その復原は正しい箇所も、現在の水準には至らない箇所も存在する。その復原に対する理解も、明治30年代における正倉院文書理解への基礎史料であるといえる。これら写本や、『大日古』のテキストのありようの分析は、近代における、正倉院文書への理解のあり方への研究、ひいては、近代初期における古代社会への理解を知ることへの一材料となるであろう。

以上が、主要な正倉院文書の分析における、「基礎要素」である。

3. 情報の統合と検索システム (2)

SOMODA の検索・および参照の総合的なシステムを図1として表示する。

参照される資料が多様であるということは、検索されるキーが非常に多様であるということでもある。そのため、検索すべきキーを、全文検索・大日本古文書・現状文書の構造に即して・そして写経事業単位といった多岐にわたるものを用意することとした。それぞれ、検索の結果において、表示する画面は同じである。検索結果より、研究者の要請に応じて、現状・復原の両様態での閲覧分岐を作っている。

また、検索結果画面のなかで、もっとも詳細なものを図2として表示する。

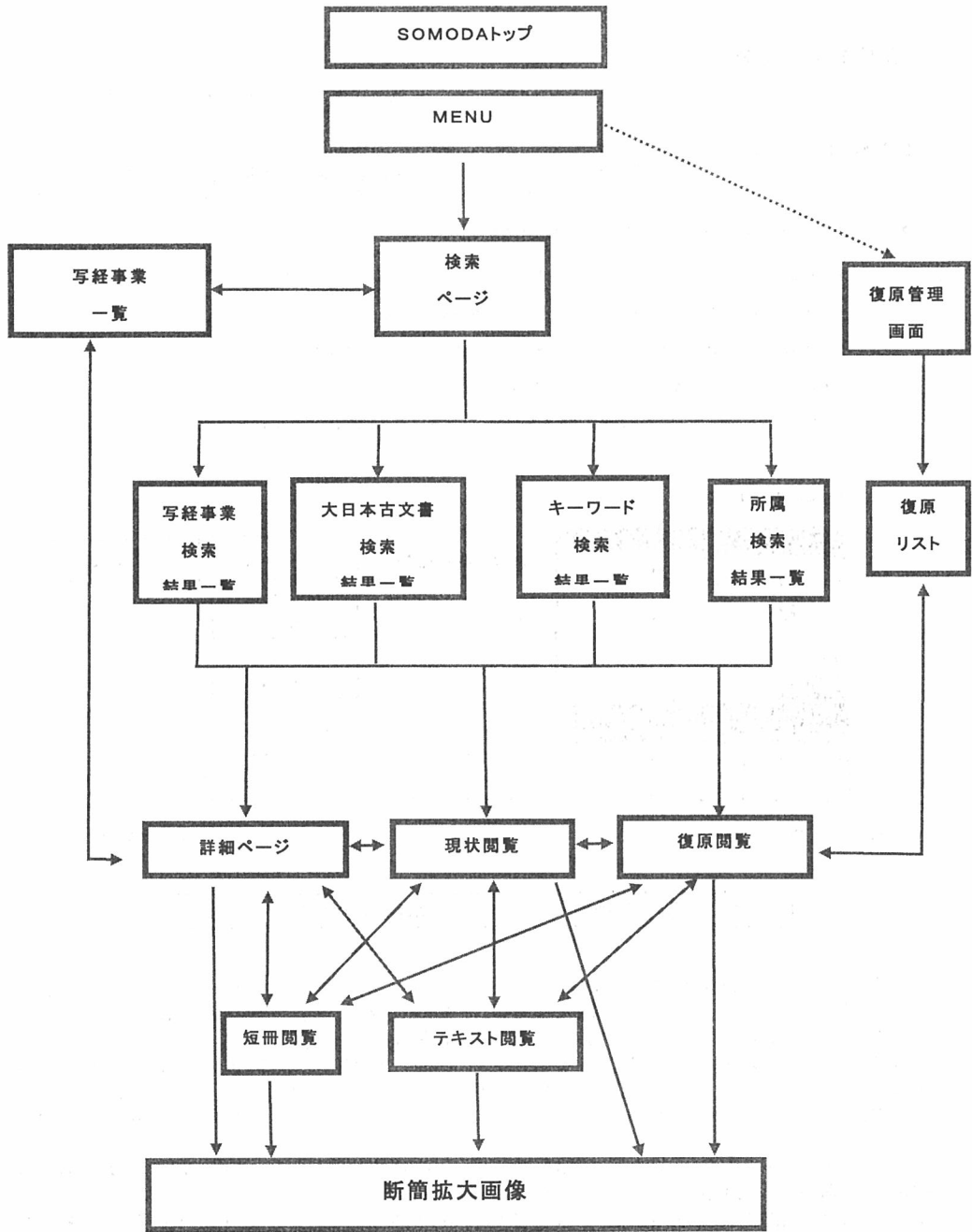


図1 SOMODAにおける検索・表示システムの実現図
 ※矢印は相互の参照関係を示す。作図の都合上、断簡拡大に収斂するように見えるが、詳細ページ・復原・現状・短冊・テキスト・拡大画像はそれぞれ研究者の要請に応じているので、すべて最終的なデータの表示である。

正倉院文書 詳細

編 正倉院文書データベース作成委員会

<< SOMODA HOME

検索画面へ戻る

検索結果一覧画面へ戻

閲覧画面へ

文書名〇〇〇 所属:正徳〇〇巻 第〇〇断 大日本古文書巻・〇〇〇

ID 〇〇〇

上政戸国造族石足戸口十三<正丁二 少丁三 緑児一 兵士二 小子二 井十 正女二 緑女一 井三>

下々戸主石足<年卅三兵士> 戸主兄国足<年卅四正丁> 嫡子安倍<年六小子>

戸主弟高嶋<年廿七兵士> 嫡子八十麻呂<年二緑児> 戸主弟久留麻呂<年廿五正丁>

次大黒<年廿少丁> 次広国<年十九少丁> 次友乎<年十八少丁>

戸主甥奈世麻呂<年十小子> 戸主母国造族麻奈亮<年四十七正女> 戸主妻国造族志邇多女<年三十二正女>

大黒児阿尼亮<年二緑女>



■ 表・裏を確認す

TEXT 表・裏

画像 表・裏

■ 写本

写本

写経一覧

カラー

モノクロ

前ページへ

次ページへ

図 2 正倉院文書データベース検索結果のサンプル画面

文書・断簡に関する詳細な情報は詳細画面として提示することとした。詳細画面では、画像のサムネイルとテキスト・料紙情報と接続・写本の有無（写本がある場合にはその史料へのリンク）・写経所文書の場合には関連する写経事業などの情報を提供する。復原においては、可能なかぎり、文書の上下端のサイズなどを正確に表現し、原文書の様子を詳細に示すよう工夫している。また、煩雑なテキストを読解可能なように、ALT タグを用い、マウスオンによる判読機能をつけることとした。このことにより、テキストと画像の併用を可能とした。

これにより、基本的な情報を統合して閲覧することが可能となっている。また、接続関係などの情報は短冊形式での閲覧を可能としている。

しかし、接続情報など以外の多種多様な情報は詳細画面においては閲覧可能であるものの、文書全体にわたってという形式では復原・現状にせよ一覧できる状況にない。このシステムの開発が今後の課題といえる。

4. 情報の統合と検索システム (3)

本システムによって、2. で述べた要素をそれぞれ表示している。それぞれの要素に対しての表示方法と、歴史的な分析の可能性を述べる。

(1) テキスト情報。詳細に述べる必要はないと思われる。特に検索により、大量の文書の中から、新たな発見が行われるという、「研究のきっかけ」になることが主眼となろう。

(2) 画像情報。SOMODA は、現時点では、権利の関係上、マイクロフィルムからデジタル化した画像を表示している。マイクロフィルムは昭和 30 年代に宮内庁が撮影したものが、現在の水準で見るとかぎりにおいては、良質の画像とは言い切れない。朱や、緑青といった墨ではない表記が見えない箇所も存在する。しかし、文字・切封といった、情報は十分に分析可能である。また、見えない文字や色などの情報に関しては、料紙情報で補完して対応することとした。付箋も墨で書かれているものは、ほぼ問題なく閲覧できる。

(3) 目録情報。文書の詳細な情報はもちろん、一次面・二次面を自由に移動できるようにした。一次面と二次面を同時に閲覧することにより、文書の裏表の記載関係がより明確となり、文書・帳簿の構造が分析を可能とする。この方法により、帳簿のありようがより明確にわかることとなった。

(4) 料紙のありかた。正倉院文書に関しては、宮内庁より出されている、『正倉院文書影印集成』がもっとも詳細な料紙情報を提供している。SOMODA は、この『正倉院古文書影印集成』の料紙情報を、掲載している。そのため文書の訂正関係や料紙の状況、紙質や界線にいたるまで、相当に詳細な情報を、画像、該当のテキスト（文書に訂正がある場合には、訂正前のテキストと、訂正後のテキストを同時に表記）とともに表示することを可能としている。

(5) テキストの生成について。現在、宮内庁書陵部に所蔵されている写本のデジタル化について検討中である。これらの写本と、原文書を同時に表記することにより、同時閲覧を可能とする。また、現在、使用されている、目録情報以外に、明治初年・明治中期・昭和初期の目録を収録することとした。このことにより、明治～昭和における「正倉院文書の成立」の分析も可能となってきた。また、(1) の画像情報で記した、付箋は、続々修編成時、つまり、明治中期に付された付箋であることが知られている。この付箋と、明治中期の目録との関係が深いことも知られている。この付箋と目録とを対照することも、「正倉院文書の成立」分析と深く関係するであろう。また、写本のテキストと、原文書のテキスト、『大日古』の差分を分析することにより、『大日古』が参照した写本類の分析や理解も進むであろう。

5. 文書復原閲覧

SOMODA の主要な目的として、上記に述べた一覧性の高さと同時に、復原表示の問題がある。復原に関しても詳細なインターフェースが固まりつつあるので、ここに報告する。

SOMODA における復原の表示は、図 1 にも述べたように、短冊と復原閲覧画面とがある。短冊は、帳簿全体を表裏同時表記でみるものである。サムネイルを表裏同時に表記することにより、帳簿の構造全体を概観することとなる。ここでも、接続情報などの概観も可能である。また、サムネイルをクリックすることにより、3. ～4. で示した、詳細画面へとリンクしている。

復原閲覧画面は図 3 のとおりである。指定した紙を基準に、復原閲覧と現状閲覧画面の両方を自由に行き来できるようにし、現状と復原の比較を行えるようにした。これは、当初から期待されたもっとも基本的なインターフェースであるといえる。このインターフェースを実現するために、データベースの元画像を 25×25pix 程度のグリッドに分割することとした。分割した理由は、1. 復元画面における、両者の文書の間をより明確に示すため。2. Alt タグにより、写真上にテキストを表示させるため。3. 単純なダウンロードによる複製の防止。の 3 点である。3. の画像の複製防止は、史料の性質上可能なかぎり防ぐ必要がある。画像を細かく分割することにより、右クリックなどによる単純な画像の複製を防止することとした。

6. 写経所文書研究会データベースとの連携

SOMODA は、単純な史料そのものの情報のみならず、研究の成果による、データの分類を可能とすることとした。それが、「関連文献目録」と、「写経所文書研究会データベースとの連携（組み込み）」である。まずは、「写経所文書研究会データベースとの連携」について述べたい。

正倉院文書は全体としては漠然とは性質が理解されているものの、「何のために残されたのか」などといった、文書のありようは一切判明していない。まして、中の一点一点につい

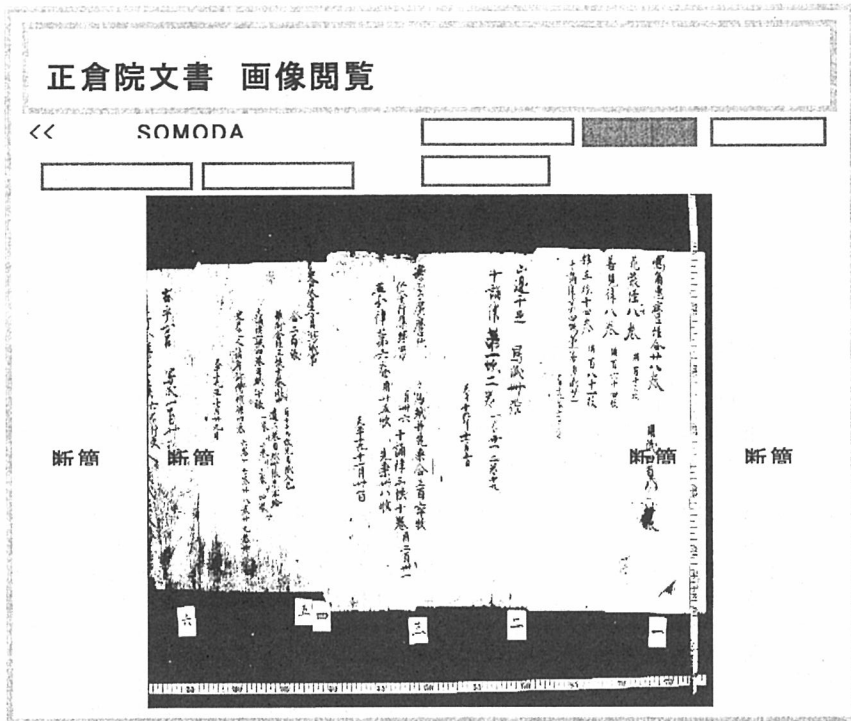


図3 復原・現状閲覧画面サンプル
メインとなる紙の左右にそれぞれ復原・現状どちらを表示するか、自由に指定できる

⑤ 研究者による並べ替えのパターン作成

研究者は管理画面で「並べ替えのパターン」データを作成することができます。

- 1 登録された研究者は管理画面での変更ができます。
- 2 登録パターンデータは研究論文とする。(巻末情報も記載する)
- 3 管理者は決められたフォームのEXCELデータを更新し、CSVデータをアップロードし復原パターンを登録・変更できる。

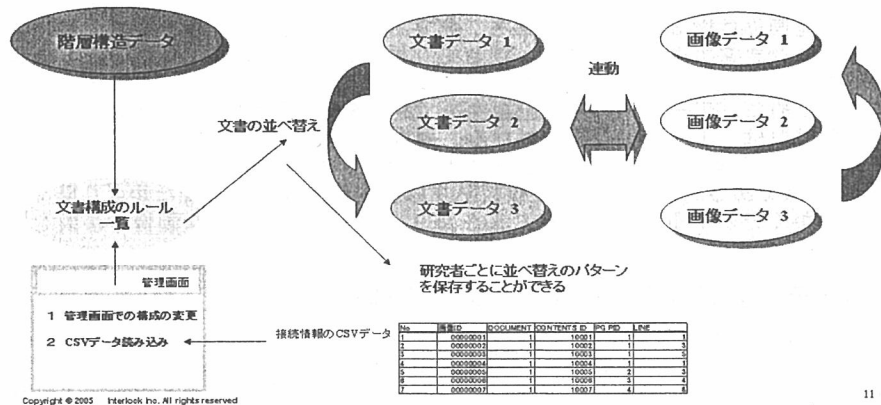
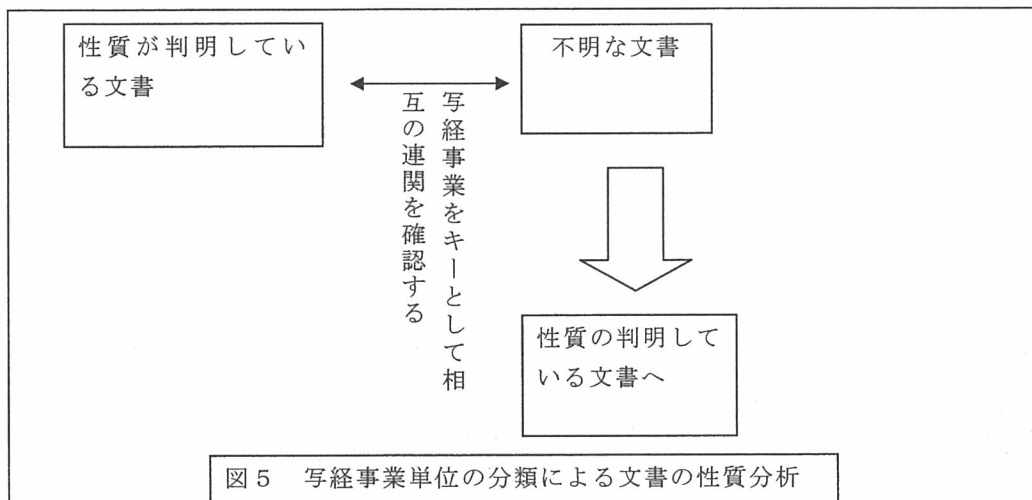


図4 研究者による正倉院文書復原



ては、まだ詳細な情報が判明していないものが多い。年紀が不明なものや、どのような事象についてかかれたものか不明なものはまだ多くある。その一点一点を明らかにし、正倉院文書全体の性質をもより詳細に明らかにするための手法として、現在とられているものの一つに、正倉院文書のなかにあらわれる写経事業の研究というものがある。

現存している正倉院文書のほとんどは、造東大寺写経所といわれる写経所の文書群と、造石山寺所の文書群である。戸籍・計帳といったいわゆる公文類は、反故として利用された結果、残ったものである。「塵芥」に若干、大仏開眼にともなう文書類が、また、ごく一部に何らかの事情で流入した文書がある。しかし、それら以外は、すべて、写経所・造石山寺所関係文書のどちらかに分類されるといってもよい。特に造東大寺司写経所の文書がその中心的なものといってもよい。

正倉院文書の中でも、造東大寺司写経所のものをここでは「写経所文書」と呼称する。写経所文書とは主として、写経所による經典書写事業、つまり写経事業の関連の帳簿である。造東大寺司写経所が、依頼された写経事業に応じて、写経を行う。そのときの事務運営として、帳簿が作成されている。

その写経事業を単位として、現存している帳簿類を分類することにより、正倉院文書の一点一点の性格を明らかにするねらいである。ある性質不明の文書があったとして、その文書がどの写経事業に関係する文書が分析することにより、帳簿や断片の性格が判明する。正倉院文書の大部分を占める写経所文書の性質と構造を理解することにより、正倉院文書のかなりの文書類の理解が進むと考えられる。また、これは、いまだに未解明である部分の文書復原においてもおおきな役割を果たすといえる。二次面である写経所文書の性質が判明することにより、八世紀における東大寺写経所の実態解明や、一次面（戸籍などの公文類）の復原や分析へとも進むこととなる。文書の大部分の性質が明らかとなれば、そもそも正倉院文書は、なぜ、東大寺正倉に収められることになったのか（＝現在まで伝来したのか）、といった分析が可能となる。無論、当時の写経所の政策や、仏教のありようもわかる。

それは、研究の間口が一挙に広がることを指名している。そのため、現在 SOMODA の主要な機能として、個別の文書がどのような写経事業に関係する文書なのかというデータベースを組み込むようにしている。写経所文書の総合的な分類自体は、すぐれて人文科学的な課題であるが、写経事業単位で検索をかけ、関連文書群を自由に表示するなどの機能においては言うまでもなく、コンピュータが圧倒的に勝るため、現在、SOMODA の主要な機能として位置づけ、作業を行っている。

「文献目録」も入力作業を行っている。当初は、関連研究文献の一覧のみを提示し、いずれは資料と研究文献との関連を目指す。

4. 研究成果データベースとしての可能性

現在のデータベースは、主として、さまざまなデータに「あたりをつける」ためのデータベースが多い。それは、コンピュータとして、もっとも重要な特徴の一つである、「全体を

見渡す」ということに由来しているといえるだろう。しかし、一方で、研究成果をまとめるための手段としてのデータベースという可能性も、模索されている。研究成果をデータベースとしてあらわすことは、人文科学としてはこれまでなされてこなかった。あたりをつけるため、研究成果として、その両者にこの SOMODA を位置づけることができればと考えている。正倉院文書の画像や、目録情報を詳細にみることのできる研究者は日本古代史を専攻するものにおいても多くはない。それは、基礎作業の困難さと同時に、閲覧環境が限られているという現状も存在する。そのため、まずは、閲覧環境の充実という基本的なデータベースの役割をも果たす必要がある。無論、文書を復原状態で提示することは、研究環境の飛躍的増大であることは言を俟たない。同時に、正倉院文書を研究するものたちにとっての成果の集積にもなることを目指している。一つは復原情報の集積、もう一つは、写経所文書の分類情報の集積である。復原情報は、登録した研究者であれば、さまざまな復原の「案」を提示することを可能とする(図4)。そのことにより、研究者による、正倉院文書の構造分析の成果を、集積することをめざす。写経所文書に関する分析も、継続的に入力続けることにより、より深い分析が可能となるだろう。特に、研究の成果によって、史料の状況が物理的にも大きく変化するという意味において、復元や、写経所文書の分類といった研究成果の表示は、コンピュータに向いているといえる。写経所文書の成果はもちろん、さまざまな研究成果を載せることにより、新たなデータベースの形を模索したい。

参考文献

- [1] 栄原永遠男：『奈良時代の写経と内裏』、2000
- [2] 栄原永遠男：『奈良時代写経史研究』、2003
- [3] 西洋子・石上英一編：『正倉院文書論集』、2005
- [4] 後藤真、柴山守：正倉院文書研究資料の XML/XSLT による記述と統合、『人文科学とコンピュータシンポジウム論文集』2002、pp.209~216、2002
- [5] 石上英一：『日本古代史料学』、1997
- [6] 後藤真、柴山守：正倉院文書復原過程の XML/XSLT による記述、『情報知識学会誌』11 巻 4 号、pp.2~16、2002
- [7] 後藤真、柴山守：正倉院文書の情報化と復原、『正倉院文書研究』、第 9 号、pp.130~149
- [8] 後藤真、柴山守：正倉院文書データベースによるあらたな研究展開の可能性とその課題、『人文科学とコンピュータシンポジウム論文集』2004、pp.51~58、2004