

## 正倉院文書データベースによる新たな研究展開の可能性と

### その課題

後藤 真<sup>†</sup> 柴山 守<sup>‡</sup>

大阪市立大学 大学院文学研究科<sup>†</sup> 京都大学 東南アジア研究所<sup>‡</sup>

本報告は、今まで報告を重ねてきた正倉院文書研究支援システムの実現形態とその状況を報告するものである。筆者らは、今まで正倉院文書をより詳細に分析し、研究するツールを作成してきた。本文、写真（マイクロフィルム）、目録をXML/XSLTとブラウザにより同時に表示することにより、文書の分析にかかわる諸情報を統合し、復元形態を自動的に提示することにより研究を支援する仕組みである。筆者らはこれを正倉院文書研究支援システムと呼称し、開発を続けてきた。今年度より、同システムは正倉院文書にまつわるより多くの知識をデータベースとして導入した新たなデータベースとして開発を進めている。筆者らはこのシステムを正倉院文書データベース（SOMODA）という名前とし、より高度なデータベースとした。本報告はそのデータベースが人文科学研究、特に歴史学研究にいかなる意義をもたらすか、また、データベース化に際しての一般的な課題と正倉院文書特有の課題を検討したものである。

## The possibility of new research deployment and issue using Shoso-in Monjo

### Database (SOMODA)

Makoto Goto<sup>†</sup>

Osaka City University<sup>†</sup>

Mamoru Shibayama<sup>‡</sup>

Kyoto University<sup>‡</sup>

We report the realization form and situation of the Shoso-in document research support system by which the report was come in piles until now. We have created the tool which analyzes the Shoso-in document in detail, and is studied. It is the structure which supports research by displaying texts, photograph (microfilm), and catalogues simultaneously by XML/XSLT and the browser by unifying many information in connection with analysis of documents, and showing a restoration form automatically. We named this system the Shoso-in Monjo research support system, and have continued development. From the current fiscal year, it is furthering development, using this system as the new database which introduced many knowledge as a database rather than it coils round the Shoso-in document. We made this system the name of the Shoso-in document database (SOMODA), and were taken as the more advanced database. This report examines a general subject with, and a subject peculiar to the Shoso-in document.

#### はじめに

筆者らは、これまで八世紀日本における一次史料である正倉院文書の復元を目指して、総合的なデータベースシステム作成を行ってきた。正倉院文書は、その歴史過程の中で、一次面と二次面に混乱関係をきたし、さらに、

断簡単位に分割されるという状況となった。

そこで、コンピュータ上で復元に関する諸情報を統合し、表現することによって、コンピュータ上での復元を目指しシステムの理論的な検討を行ってきた[1][2][3]。

上記の復元システムは、マイクロフィルム画

像(D1)、目録(D2)、翻刻(D3)を連動させて、同時に表示させるシステムとして作成している。このシステムを筆者らは今まで正倉院文書研究支援システムと呼称してきた。正倉院文書研究支援システムは、XMLによる詳細なタグ付けにより、正倉院文書を復原させるさまざまな情報を表示させ、復原させることを可能としたシステムである。このシステムの開発により、日本古代史における写経所の研究をはじめとして、古代社会や経済の構造、官司制度の構造や、日本古代仏教政策の構造などの解明に寄与することを目的とした。

## 1. 正倉院文書データベース(SOMODA)

今年度より、この研究支援システムの実現環境が整ってきた。『大日本古文書』のデータ、東京大学史料編纂所作成の『正倉院文書目録』に関しては東京大学史料編纂所のご協力を得て、編纂所データベースの『大日本古文書データベース』を援用させていただくこととなった。また、マイクロフィルムの画像データに関しては宮内庁正倉院事務所より学術研究に限りご許可をいただいた。さらに、正倉院より流出した古文書類に関しては、国立歴史民俗博物館の協力をいただけることとなった。これら関係各所のご協力により、各データを蓄積し、公開できる環境を整えることができた。

この新たなシステム作成に伴い、より多くの知識を導入し、高度な研究用データベースとして作成することとなった。そのため、正倉院文書研究支援システムは、正倉院文書データベース(SOMODA)として、衣替えを行い、開発を行っている。SOMODAは正倉院文書データベース作成委員会を作成し、正倉院文書研究に深くかかわっている関係の皆様が委員として参加していただいた。その委員の皆様のご協力を得てシステムを作成している。

SOMODAは正倉院文書研究支援システムの上に、さらに多くの知識を加えた総合データベースシステムである。SOMODAは、明治期の目録・写経事業単位での帳簿の集積知識・人名辞書や、関連論文データベースなど

さまざまな情報を加えることとなった。正倉院文書研究支援システムにおいては、特に正倉院文書の中心をなす、写経所、および造石山寺所といわれる官司群の帳簿の復原が可能なシステム構成が中心であった。

SOMODAでは、復原にかかわる情報に加えて、正倉院文書にかかわるより詳細な情報を加えたこととなる。(図1および図2)とくに写経所事業単位でのデータベースによる文書群の索引は正倉院文書独自の特徴を現したデータベースシステムであるといえる。

## 2. 正倉院文書XMLの基礎構造

正倉院文書をXML化するためには、史料の階層構造を把握する必要がある。正倉院文書は明治期の整理を経た現状は、史料名の総体としての正倉院文書、史料群名としての写経所文書(続々修の場合には帙)、巻、断簡、紙の階層構造をなしているといつてよい。通常の史料用語においては、断簡とは文書の一断片を定義する。しかし、正倉院文書においては、奈良時代の現状を維持している、一枚ないし数枚の紙の集合を「断簡」と定義し、呼称している。これは、正倉院文書が経てきた、特殊な事情を示すと同時に、正倉院文書の構造をもっとも端的に表しているものといえる。

したがって、正倉院文書を史料群体や、構造として見る場合には、帙、巻、紙などの情報とともに断簡として見るのが肝要である。断簡単位での処理は、『正倉院文書目録』などでも一般的に行われていることである。直接の研究ではみえてはこないが、正倉院文書になんらかの類型処理を加える場合には基本となる考え方である。(以下では、この処理方法を断簡分類主義と定義する)

この断簡分類主義に基づき、SOMODAでは、各断簡すべてに識別番号であるIDを付することとした。IDとタグ付与の基本的な考え方は以下のとおりである。

1. タグが<正倉院文書><所属><帙><巻><断簡>が構造記述の基本である。
- 2.1. をふまえ、大日本古文書を識別する<大日本古文書 id・巻・頁>を挿入して、断簡と大日本古文書の二重の構造表記を可能とする。

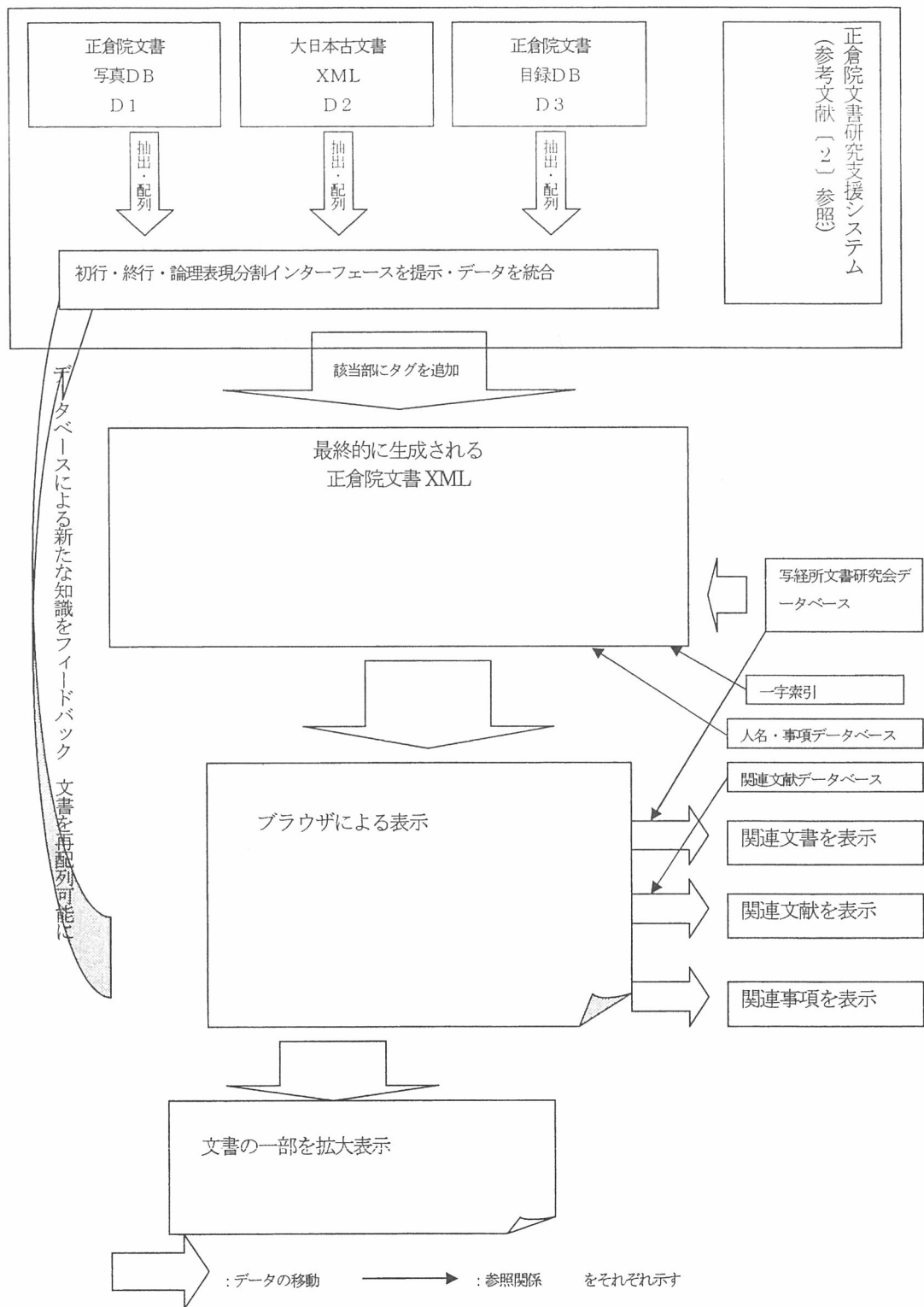


図2 SOMODA 概念図

3. 割注や訂正などを示すタグとして、その他雑多なものを定義した。最終的に生成された XML 文書には、紙、行の情報もタグとして挿入されている。(本来ならば、この紙、行も史料構造の一つであることは疑いの余地はない。しかしながら、抽出・再配列などの処理の最小単位はやはり断簡であることから、<紙><line>は 3. のテキストの属性であると分類することとした)

ID は先述した構造に即してつけなければならない。したがって、所属(1)+帙(2)+巻(2)+断簡(2)+表裏(1)として ID を付した(カッコ内は数字表記の桁数)。まずは、所属である。所属は正集を 1、続修を 2、続修後集を 3、続修別集を 4、塵芥を 5、続々修を 6 とする。つぎに帙に番号を付す。01~47 となる。ちなみに、続々修以外は帙という単位は存在しないのですべて 00 とした。巻は巻次順に 01~XX (最大 50) までとした。断簡も同様に断簡順に 01~XX (最大 99) までとした。現状では 1 巻の中で断簡数が 100 を超えるものはないと確認している。表裏は、表を 1、裏を 2 とした(例外の可能性は、なお残るが、現時点では考察の対象としない)。

この考え方で、ID を付していくと、たとえば続々修 18 帙 6 巻表第 1 断簡(文書自体の表現なので、句点は入れないほうがよいと思います)ならば 61806011 となる。したがって、XML 文書では「<断簡 pid=" 61806011">」と記述する。

この断簡を最小単位として、自由に処理できる単位の XML オブジェクトを生成する。



図1 正倉院文書データベースの結果表示

### 3. 正倉院文書のデータベース化への課題

本報告では、とくに正倉院文書をデジタル化する場合においてあらわれてきているさまざまな課題を列挙してみたい。SOMODA では、マイクロフィルムデータを画像データとして保存する以外はすべて XML として記述している。とくに翻刻情報を詳細に XML 化するためにはさまざまな問題が浮かび上がってきている。以下にその課題を列挙してみたい。

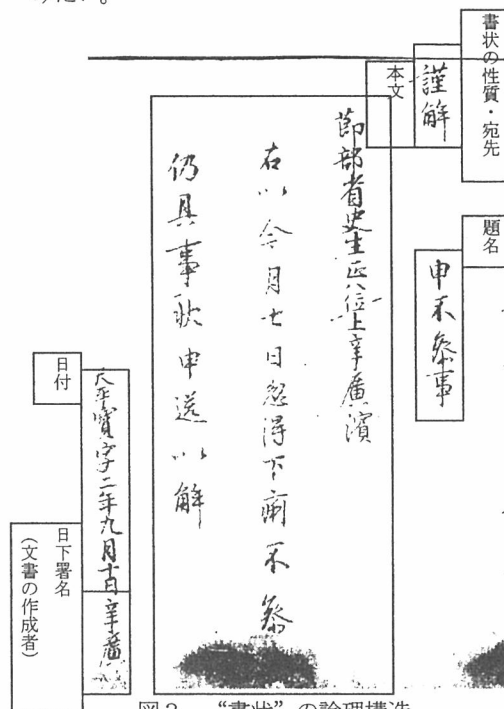


図3 “書状”の論理構造

#### 3-1 XML 化への課題

正倉院文書の本文を実際に XML として定義する場合、さまざまな課題があらわれてきている。XML では本文を一次線上のテキストで描かれることを前提としている。しかし、(前近代の文書類が全てそうであるように)正倉院文書は手書きであるがゆえに、単純な一次線表記で描かれない事例がむしろ多い。日本古代史料においては、一般的に右上から左下にかけて、縦向きにかかれ、下端に達すると、一行左に移動する、という方式をとっている。この仕組みはいわゆる書状類に関しては有効な視点であるといえる(図3参照)。書状では、発信者、受信者、文書の性質(解・

移・牒・符など)、本文、日付、日下(文書の作成者)、が明確な法(公式令)規定、および慣習により構成されている。また、一次的な書記、二次的な書記と順序が比較的容易に理解できる。このことにより、XML による表記上の構造化と、論理上の構造化が比較的容易であるという特徴がある。

しかしながらいわゆる帳簿の形態では必ずしも上記のようにはいかない現状を指摘できる。帳簿は基本的に内部文書である性質上、その帳簿作成者の都合に応じて自由に作られる。また、追記関係が交錯し、明確な論理構造が記述できない場合が一般的であるといつてよい。

帳簿を XML 化する場合には、明確な論理構造記述を放棄し、一定のクラスタ単位でタグ組みを行い、叙述する可能性を検討する必要がある(図4)。

XML は、オブジェクト指向言語において、タグの一要素をオブジェクトとしておくことを可能としている。そこで、帳簿のあるテキスト群を塊として理解し、タグ付けを行う。その上でそのタグに囲まれた一つをオブジェクトとして理解し、処理する。このことにより、正倉院文書の複雑な表記がより簡明に XML として記述できるのではないかと現在検討を進めている。

8/558-90

```

<紙 paid="12">
<text txid="1">十八年正月廿七日</text>
<text txid="2">撰大乘論疏第三</text>
<text txid="3">用</text>
<text txid="4" zoku="black">七</text>
<text txid="5">紙七十七</text>
(中略)
<husen daid="34">卅四</husen>
<husen meid="2314">廿三の十四</husen>
<紙>

```

図4 クラスタ単位での XML 処理の可能性

#### 4. 写経事業データベースと検索

検索も重要な課題である。正倉院文書は上記のような文字の通用以外にも、同一人物が別の名前で多々登場したり、同一の官司が名前を変えて登場したりすることがままある。そのために、代表的な、人名や官司名、物品名などは専用の知識データベースを作成し、そこで検索をより高度にする仕組みを作成している。ただし、知識データベースに格納する知識自体の正しさを担保するため、その知識に関する参考文献の掲載なども検討中である（図5）。

さらに、写経事業単位での帳簿群の検索を可能とするシステムを開発している。正倉院文書は、伝来の段階では主として東大寺写経所の文書群と、造石山寺所の文書群が中心であったことが判明している。そこで、東大寺写経所の写経事業単位で文書群を再構成し、復原することによって、正倉院文書がより深く理解できるのではないかと、の視覚のもと、写経事業単位のデータベースを作成している。

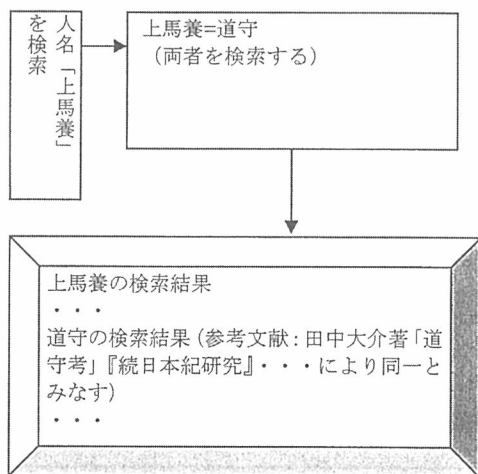


図5 知識データベースによる検索

正倉院文書の中でも、写経事業に関わる文書（＝写経所文書）のデータベース作成を一つの目論見としている。目的は、

1. いわゆる写経所文書の各断簡がどの写経事業の関連史料かを示す。
2. 写経所に関わる史料を写経事業単位ですべて分類し、正倉院文書の全体像をスケッチする。
3. 写経所のあり方をより深く研究する材料とする。以上3点である。

具体的なデータベースの内容は、大きく3つに分類できる。事業データベース、史料データベース、複合帳簿データベースの3つがそれである。

事業データベースとは、東大寺写経所で行われた写経事業を一瞥できるようにしたものである。正倉院文書にみえる写経事業の一覧は菰田香融氏〔4〕が、おおよそのものを作成している。このデータを母体として引継ぎながらも、より精細、かつ網羅的なデータを作成することも目的としている。写経事業単位のデータベース群の基礎情報である。

史料データベースは個々の写経事業ごとの史料を断簡単位で示したものである。断簡主義の考えかたは前述したとおりだが、その方式は写経事業単位のデータベースでも踏襲している。データイメージとしては、事業データベースに付随するものであるが、このデータベースの主眼である、「史料と写経事業の関連付け」という意味では、もっとも主要かつ重要なデータとなる。

史料データベースの概念は、前述したように、事業データベースに付随するものとなる。従って、複数の写経事業にまたがる文書類を処理することは困難となる。そこで作られたのが複合帳簿データベースである。

複合帳簿データベースとは、複数の写経事業が同時に記載される史料のデータ化を試みたものである。しかし、このデータベース作成には困難が伴う。データの項目は史料ごとに規定されてしまうからである。たとえば日次式の帳簿と、口座式の帳簿、この両者をとってみても書かれている項目も異なるし、また、手実帳のような特殊な帳簿形態もある。これらを厳密にデータ化し、整理することはすべての帳簿の性質の検討をすませた後となるであろう。

データ入力最大の眼目は、ある断簡と写経事業とを関連付けることである。そのためには、断簡分類主義をとることがもっとも明瞭なデータ整理となるであろう。たとえば、史料データベースでは史料の詳細な情報を記載するのではなく、該当写経事業がどの断簡にあるのかを示すことが重要である。不必要なまでの史料の詳細な情報は、データを煩雑にするばかりでなく、ユーザのデータの見方を規定してしまいかねない危険をはらむ。その危険を回避するためにも、史料の配列などは断簡単位で処理する必要がある。ただし、その場合には続々修の断簡をどのように認識

するかという問題は残る。

それでは、この写経所文書研究会データベースの意義はどこにあるのであろうか。このデータベースの目的は、ある個別写経事業に関わる史料を整理するという、同様にコインの裏表の関係として、ある断簡がどの写経事業に該当するのかを明らかにするということである。それは、非常に多数の写経所文書の中で、今まで不明確であった史料の位置づけを明らかにしようとするものである。大量にある写経所文書の中には、いまだにその史料の位置づけが不明確なものもある。そのような文書をすべて位置づけようという意図でもある。また、それは全体の復原作業の一環ともなる。

### 5. 復原の方法とテキストの三次元性

いわゆる文書類のようなテキスト群は見た目上は平面状に二次元でテキストが羅列されているように見える。しかしながら、実際は文書の貼りつき面や、付箋の情報などは三次元的に構成されているのである〔5〕(図6)。そのため、文書の復原にかかわる情報では、三次元的な文書の貼りつき面が重要な要素となる。そこで、SOMODAでは、文書を三次元的に理解し、処理可能なように復原システムの作成を考えている。特にテキストのみでの復原ではなく、画像データを自由に貼りかえることにより、墨痕やしみ、紙の切れ目などによる連続と断絶を確認できるシステムとして作成を試みている。しかしながら、写真の情報にも限界がある。また、マイクロフィルム自体が比較的古い時期に作成されたこともあり、デジタル化にむけてコマのサイズが一般的でなく、広い。濃度ムラがひどい。カスレ等が多いなどの状況がある。そのため、写真の情報の改善など、より復原に必要なシステムへと改善の余地がある。

### 6. 正倉院文書における古代史研究上の意義

SOMODAによる、正倉院文書の詳細な復原と研究情報の提供により、多様な日本古代史研究上の課題解決への支援となることが予測される。復原への支援となることはもちろん、なによりも、古代帳簿および文書によるシステムの解析などが、まず最初にあげられるであろう。XMLによる詳細な帳簿の構造化は、研究者がいままで気づかなかったであろう帳簿の詳細な情報を提供することとなる。

また、写経事業単位による帳簿の分類が可能となることにより、正倉院文書がなぜ、正倉院に残されることとなったのか、が解明されるものと筆者らは考えている、正倉院文書が残った最大の理由は、反故として、物品を包んだりするものであろうという説〔6〕と、誰かが何らかの意図をもって持ち込んだのではないかという説がある。もし、写経事業単位によるデータベース構築により、文書の残りようがある一定のものであったことが判明したならば、それは正倉院文書のみならず、当時の政治状況や、社会状況の解明に重要な役割を果たすであろう。

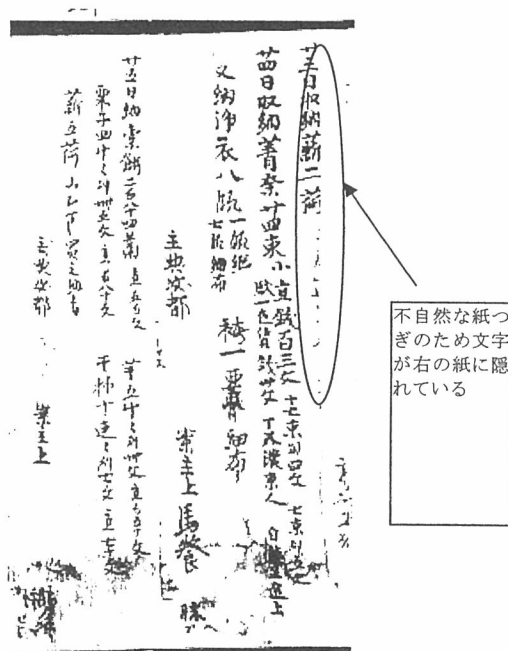


図6 文書の三次元構造

### 7. 総括と課題

今までの正倉院文書研究支援システムにおいては、正倉院文書に関わる本文情報と同時に、画像情報、目録情報を表示させる総合的データベースシステムであると同時に、天保～明治期にバラバラにされた文書をバーチャルで復原させることが可能であるという点に特徴があった。SOMODAは、この基本的属性に加え、正倉院文書に関わるあらゆる基礎データを付加することにより、より高度な研究支援システムへと昇華させたものである。そのSOMODAの基本システムは図1の通りである。基本的な骨格は、支援システムのものを用いるが、以下のように新たな機能を加

えた。1. 写経事業にもとづくデータベースを参照することにより、文書の復元だけではなく、関係文書を同時に探し出すことを新たに可能とした。2. 検索・一字索引機能、詳細な料紙情報、明治の目録などを加えることにより、より詳細な正倉院文書に関わる研究の情報を提示する。3. 該当資料にかかわる研究論文を表示し、研究史の整理をも可能とした。4. 上記の情報を研究者が参照し、正倉院文書の“断簡”を自由度高く再構成できるように、システムを改良した。このことにより、今後への新たな研究展開の可能性を大きく残し、正倉院文書をより広い研究ツールとなるものである。たとえば、文書様式の変化の分析は、古代古文書学に大きな影響を与えるであろうし、古代帳簿学の構築の基礎となるであろう。また、写経事業の比較分析が容易になることにより、国家仏教の動向を探ることも可能となる。

また、正倉院文書の残り方の解明は当時の政治史や、社会史に大きな影響を与えるものである。このデータベース作成により、いまだに“使いにくい”歴史資料の一つである、正倉院文書の分析が可能となる。

しかしながら、詳細なデジタル化に向けては、XML化の方法論や、文字の問題などクリアすべき課題が多々あることも判明している。また、やや卑近な話ではあるが、XMLとして構造化することが可能な正倉院文書を理解した人材が少ないことも課題としてあげられる。

前近代の手書き文書をデジタルデータとして記述し、格納するためには、クリアしなければならない課題が多々ある。単純に“あたりをつける”ためだけのテキストデータから、研究に寄与するためのデータベースへと昇華するための一事例となれば幸いである。

#### 謝 辞

なお、本論文作成にあたり、柴原永遠男氏、山口英男氏をはじめとする正倉院文書データベース作成委員会の皆様、写経所文書研究会の皆様、東京大学史料編纂所の皆様、および宮内庁正倉院事務所の皆様より多大なるご教示をいただいた。ここに末筆ながら謝意を表す次第である。

#### 参考文献

〔1〕後藤真、柴山守：正倉院文書復元過程のXML/XSLTによる記述、『情報知識学会誌』11巻4号、pp.2～16、2002

〔2〕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

〔3〕後藤真、柴山守：正倉院文書の情報化と復元、正倉院文書研究、第9号、pp.130～149

〔4〕菡田香融：「南都仏教における救済の論理（序説）―問写経の研究―」（日本宗教史研究会編、『救済とその論理』日本宗教史研究4、法蔵館、1974）

〔5〕石上英一：『日本古代史科学』、東京大学出版会、1997

〔6〕西洋子：『正倉院文書整理過程の研究』（吉川弘文館、2002年）、杉本一樹『日本古代文書の研究』（吉川弘文館、2001年）など

#### 付 記

なお、本研究は、学術振興会特別研究員奨励費の補助を受けて執筆したものである。