

正倉院文書のデジタル化の意義と課題

SOMODAの改善データベース作成経過に即して

Significance and Issues of Digitizing “Shoso-in Monjo” :
A Study of the Improvement Process of the Shoso-in Monjo Database “SOMODA”

GOTO Makoto

後藤 真

はじめに

本稿では、正倉院文書のデジタル化の事例について過去のものから現在開発中のものまでの概観を述べ、その事例をもとに、正倉院文書のデジタル化が可能とすることについて、整理するものである。

そもそも正倉院文書をはじめとする日本古代の史料のデジタル化については、史料全体に対する比率としては多いものの、六国史をはじめとする基礎的史料のデジタル化が進められていないなどの課題を抱えている。そのような状況ではあるが、一次史料である正倉院文書と木簡についてはデジタル化が進みつつある。本稿では、正倉院文書のデジタル化の中では早い事例である正倉院文書データベース（SOMODA）の概要を述べ、その後の課題について整理し、結果として作成した「正倉院文書トピックマップ」について述べる⁽¹⁾。そして、それらデジタル化がもたらす意味について述べる。

1. 正倉院文書データベース（SOMODA）

筆者は、2002年より、正倉院文書の復元過程のデジタル化手法について、研究を行った⁽²⁾。その成果は、2004年度より栄原永遠男氏を中心とした科学研究費（研究成果公開促進費）にて入力を行い2007年に正倉院文書データベース（SOMODA）として公開を開始した⁽³⁾。

正倉院文書はその歴史的経緯により、奈良時代当時の元の状態（原状）からばらばらにされてしまった。正倉院文書を用いた研究を行うためには、この原状への「復元」を行うことが必要となる。この復元については、『正倉院文書目録』の復元情報を利用するのが一般的である。復元については『正倉院文書目録』の情報を利用しつつ、本文情報は『大日本古文書』編年1～25巻、画像はマイクロフィルムおよび『正倉院古文書影印集成』などを参照し、検討を行う⁽⁴⁾。これらの情報が分散していることも、正倉院文書研究を妨げる重要な要因となっていた。

そこで、SOMODAは上記問題をコンピュータで解決すべく、正倉院文書をバーチャルに復元し、かつ分散している正倉院文書の情報を総合的に網羅し、同時に表示させることを目指したデータ

ベースであった。

あわせて、正倉院文書の 1 紙を以下の 6 つのフェーズ（図 1）として処理し、そのフェーズを自由⁽⁶⁾に移動できることで、総合的な一つのシステムとして構築することを可能にした。

具体的には、正倉院文書のテキストデータに対して、XML（eXtensible Markup Language）によるマークアップを行い、構造化すること、画像情報や目録情報と関連付けることによって、正倉院文書のあらゆる情報を得ることが可能なシステムとした点が特徴である。

正倉院文書の一紙	
1. 正倉院文書の現状 表	4. 正倉院文書の現状 裏
2. 正倉院文書二次面 裏（奈良時代再利用面）	5. 正倉院文書二次面 表（奈良時代再利用面）
3. 正倉院文書一次面 表 （奈良時代に最初に用いられた面）	6. 正倉院文書一次面 裏 （奈良時代に最初に用いられた面）

図 1 正倉院文書の紙を左右に「開いて」みたイメージ

数字がそれぞれのフェーズ。
同じ紙の面をみても、どの時代のどの段階を主とするかで、接続関係が変わる。

これらのデータベースは、単なるテキストの全文表示や画像情報の提示ではなく、復原の情報はじめとする複数の情報を同時に持ち、かつ、それらを同時に閲覧するシステムである。正倉院文書データベース作成委員会によるデータ入力によって最終的にこれを実現した⁽⁷⁾。

SOMODA は、本文情報であるテキストデータ（『大日本古文書』編年 1～25 巻）・復原のための目録データ・マイクロフィルムを中心とした画像データの三つを大きな柱とし、これらを相互に結びつけ、かつ、復原と現状の両者を自由に配列し、閲覧することを可能としたシステムである。正倉院文書一枚の紙を図 1 のように、1. 現状表 2. 現状裏 3. 二次面による復原表 4. 二次面による復原裏 5. 一次面による復原表 6. 一次面による復原裏 の 6 つのフェーズによって表現可能なモデルの設計を行った。これらのフェーズを「断簡」（奈良時代からのままとされる紙のままとまり⁽⁸⁾）。正倉院文書復原の際には断簡ごとに接続関係を特定する）単位で自由に行き来できるようなくみとし、正倉院文書の現状・一次面・二次面の状態を閲覧することを可能としている。これは、デジタルデータの配列自由性を十分に活用したしくみであった。

紙の集合としての「断簡」という概念を用い、それにより史料を構造化することによって、復原をデジタルデータでも可能とした。復原に加え、このデータベースの作成によって、複雑かつ多様な正倉院文書の情報が、横断的、かつ総覧的に閲覧可能となり、正倉院文書の研究への寄与を果たすことを可能としたのである。

正倉院文書については、そもそも専門の研究者が使いたいというものから、「ちょっとどのようなものがあるか見てみたい」という他分野の研究者の要望までさまざまなものがある。その中では、専門の研究のツールという側面が強く、正倉院文書をちょっと見たいユーザには使いにくい部分があるとはいえ、実際にデータを入力し、共有を推進するという意味においては、重要な試みであったと位置づけることができる。

2. SOMODAの課題

1で述べたようなSOMODAが目指したデータベースのモデル（断簡を復原の基礎とする断簡単位主義など）に関しては、一定の評価を得たと自負しているが、最終的なその実装システムについては、さまざまな課題を残した。具体的にはインターフェースの不完全さと、検索システムと画像表示の不具合である。小口雅史氏は、そのシステムの「重さ」なども含めて、今後の課題が多いことを指摘している⁽⁹⁾。また、データベースが正倉院文書という史料の構造に即したものになっていないのではないか、という疑問を投げかけられている⁽¹⁰⁾。

これらの批判は、SOMODAを「どのようなデータベース」として位置づけるかとも関わってきている。復原などの研究ツールとしての目的を優先させるのか、それとも、基礎的な情報共有を優先させたものなのかのバランスの問題である。作成当時の状況においては、正倉院文書という史料単位でのデジタル化を、実際の史料保存機関や編纂に関わる機関ではない、大学という研究機関で行うことの位置づけが、はっきりとしていなかったこともある。そのため、まずは、データの作成のみが優先され、結果的に研究・解析に用いるデータベースか、基礎情報の提供か、という整理が曖昧であった部分はたしかにあった。

そして、システムのモデル自体は断簡を最小単位に復原のデータを構成するという断簡単位主義をとってはいるものの、本文データの構造が断簡単位主義ではないという矛盾を結果的に抱えており⁽¹¹⁾、この矛盾を解消するために多くの無理を重ねていた。たとえば、現時点でのSOMODAのテキストデータは紙単位でタグを付し、それがRDBに格納されている。断簡と紙との対応関係には、専用のテーブルを作成しているなど多くのデータを同時に参照する必要がある。また、断簡単位でデータを処理する際には、クリックするたびに常にその対応テーブルを参照するという、処理回数の多いデータ構成となっている。その結果、このようなシステムの関係上、データへのリンク・共有なども困難なものとなっている。これらの問題点を解決することが、SOMODAの「次」の重要な課題となっている。

これらの問題を解決するためには、断簡単位主義の徹底を目指したデータ構造の修正を図らなければならない。さらに、Linked Data⁽¹²⁾の隆盛に見られるような、近年のWeb上での効果的なデータの相互参照の流れを考えるならば、可能な限りシンプルな構造のデータとし、かつ、永続的なリンクを貼れるようなシステムを構築することを考える必要がある。

3. SOMODAの課題解決と今後

そこで、SOMODAのデータを改良し、以下のようなデータベースの作成を行っている。SOMODAの第2版として位置づけ、正倉院文書の情報をより高度に利用できることを可能とするシステムを目指している。システムとしては、トピックマップと呼ばれる、概念語をネットワークで結びつける手法を採用している⁽¹³⁾。トピックマップはオントロジ技術を応用しており、多様なデータ表現と検索が可能である。

具体的には、以下のような行程でのデータ作成を行っている。本システムはSOMODAの後継と位置づけられるものであるが、両者を区別するために、以下では、新しいシステムを仮に「正倉

究論文などでの参照を可能とする形式とした。

この資料の現状構造を基本とした典拠情報群が、正倉院文書のトピックマップ全体の基礎情報となっており、オントロジ図の左部を構成している（図4）。

(2) 正倉院文書に関する研究知識情報の付与

基礎情報を付加した次のステップとして、正倉院文書に関連する知識情報を整理しトピックとして史料と結びつけた。

とりわけ写経所文書に注目し、写経所に関連する情報について記述することを目指した。具体的には、人物情報と経典情報を中心にデータ作成を開始し、関連する研究情報を付与している。

奈良時代の人名に関する情報のうち、正倉院文書に出てくるものを文献類から抽出した。そのうえで、写経所に関する研究成果をもとに、筆者が人名に対し、「身分」と「役職」そして居所について関係づけを行った。「身分」と「役職」にわけたのは、ある人物に関して、時間変化とともに変わる情報（役職）と、時間変化とともに変わらない情報（身分）に切り分ける必要があったためである。

そして、これらの人名情報を、先述の史料個別の識別情報である「紙」と関連づけた。これにより、正倉院文書に記述された人についての典拠情報が明示される。この「紙」との関連付けの特徴は、今までのデータベースのような単純に人名を文字列での発見だけでなく、同一人物の別名や、略称などが同時に発見可能になっている点にある。また、同一役職の人物を同時に発見し、原史料にあたることができるようになり、より複雑な検索が可能となる。

知識情報の付加の第2ステップとして、奈良時代に存在したことが、史料上明確な経典に関する情報を既存の文献類から抽出した。経典に関しては、それぞれに「作者」を付し、また、判明しているものについては、大正新脩大蔵経データベース（SAT）⁽¹⁷⁾へのリンクをはった。この関連付けには、現時点では経典名の一致による単純なマッチングを行っている。

これにより、大正新脩大蔵経の番号から奈良時代の経典一覧を探すことが可能となっている。現在、経典数5940（同一経典の別名有）に対し、992の経典が大正新脩大蔵経の経典名と一致したため番号を付している。同一経典名で作者（訳者）とテキストの中身が異なる場合、（わかる場合は）人間が作者・巻数等を同定した上で、人間の手でIDを振りなおした。

しかし、経典名だけで正倉院文書の中に出てくる経典をトピックマップで記述することは困難であった。なぜなら、コンテンツとして記載される経典名と、史料上で出てくる実物とでは扱いが異なるためである。そこで、コンテンツとしての「経典名」と、実際に史料中に出てくる「モノ」としての「経典」を分離し、それぞれに関連付けを行った。

さらに、経典情報には当時の史料に基づいた「宗派」の情報を付与した。正倉院文書の経典目録の中には、当時の宗派ごとに経典を分類したものが残されている⁽¹⁸⁾。そのため、この宗派ごとに経典が所蔵された史料の情報をを用いて、経典のトピックに対して「所属する」という関係性を持たせた「宗派」というトピックも作成した。

また、正倉院文書ではないが、平安初期の経典目録（『五宗録』）も同様に入力を進めており、この両者を比較することで、宗派に属する経典の時間的変遷を見ることが可能になる。

4. 正倉院文書トピックマップが可能にしたこと

現段階で正倉院文書トピックマップは、具体的に以下のような有用性を持っている。

(1) 正倉院文書の人名と經典名の結びつけ

写経所において、ある人物が、特異な經典に関わっていたことがあるのかなど、通常の検索モデルだけではわからないセマンティックな情報の発見を可能とした。

たとえば図5は、作成したトピックマップをもとに奈良時代のある時点での疏の保持者とその所在地の関係を Visualize したものである。これにより、ある僧が持っていた注釈書群がどのようなものか、また、その注釈書の作者に偏りがあるのかなどがわかる。

また、史料上判明した僧の居所についてもトピックマップに入っているの、これにより、寺院—經典—僧の関係性を、經典の所持という観点から明らかにできる可能性がある。

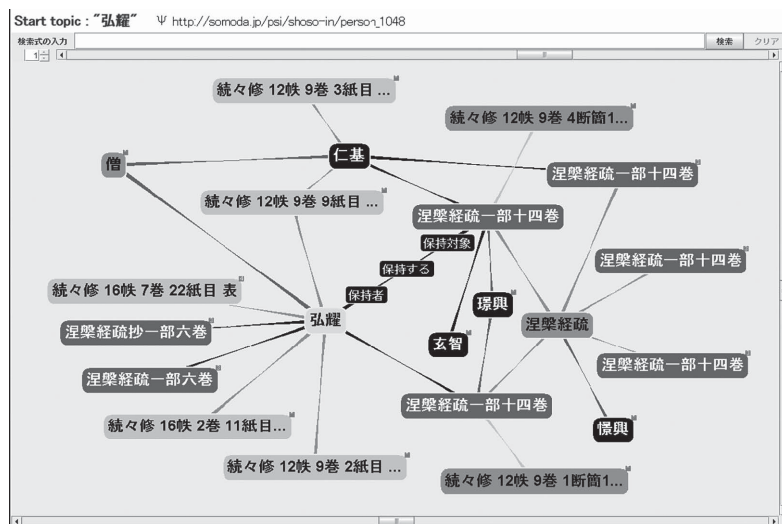


図5 僧名から經典名と個別經典・史料名の関連付けを Visualize したもの（カラーにて別途掲載）

以下のように定義されている（設定で変更可能）

- 明るいノード（上記の例では弘耀）：焦点のあたっているノード、黒：人名ノード
- 濃いグレー：經典名ノード、グレー：断簡ノード、薄いグレー：紙ノード
- 線は関連型を示し、マウスオンで具体的な関連の内容を示している。上記の例では「弘耀」と「涅槃經疏一部十四卷」の関連の線上にマウスカーソルがあり、両者の関連を示している。

(2) 帳簿と經典・人名の結びつけ

正倉院文書は、その歴史的経緯から複雑な構造を持っており、それが正倉院文書の全体像を見えにくくし、研究を困難にしてきた背景があることは前述の通りである。それを解決したのが SOMODA であった。しかし、SOMODA は、一つの帳簿を復元し、その帳簿の状況を概観するのは優位であるものの、特定のキーワードとそれに関係する帳簿を横断的に複数見るなどには向かない点があった。正倉院文書トピックマップでは、文書の復元を解決すると同時に、それらの問題点を解決できる。

特定の經典が、ある特殊な帳簿に出てくる可能性があるのか、また、ある人名の集団が、特定の帳簿に出てくるなどの傾向があるのかの分析は、奈良時代の実態に迫ると同時に、当時の帳簿の記述形式や手法といった、正倉院文書の史料論に結び付けられる。

たとえば、造東大寺司写経所には、東堂・西堂、もしくは北堂・南堂といった二つの堂舎によって作業が行われていたことが指摘されている。帳簿ごとに人名集団の傾向が判明すれば、このような実態の研究に迫ることが可能になると考えられる。

(3) 經典情報の多様化

經典の情報の多様化も可能となった。正倉院文書の中心は写経所文書であるが、仏教学などからの研究は、決して多くの量があるわけではない。複雑な経緯を持つ正倉院文書の史料的な特性がその大きな原因を占めているが、同時にそもそも奈良時代の仏教の実態を知る史料として知られていない部分もある。そのため、仏教学で多く用いられるデータベースとの連携を行うことと、經典情報を有機的に連結させ、また、復原帳簿と同時に表示することで、断片的ではない情報の提供を可能とした。

あわせて經典に対する宗派情報の付与を行っている。これにより、ある人物がもつ經典群が、ある種の宗派に偏る可能性はないのかなどをデータ上で確認できたり、日本古代を対象とした仏教学で考えられている經典と宗派の関係などとの関係性を検討したりすることが可能となる。

また、歴史学の側からも經典から宗派を知る、宗派から經典を知るなどの今までとは異なる視点の情報を提供できる。このように、写経所文書に関する情報を多様に提供することができる。

5. 期待できる成果

正倉院文書トピックマップによって、今後は以下の点が研究支援に期待できると考えている。

1. 正倉院文書を単純な全文検索・絞り込み検索だけではなく、複数の、かつ複雑な条件を用いて検索することができる（ある人物だけを検索するのではなく、ある人物が、ある時期に持っていた經典の一覧、ある人物と同一の役職を持つ人物の一覧など）。2. 研究によって判明した結びつきが可視的に確認できる（上記事例をネットワーク図で閲覧するなど）。3. PSIの機能によって、URIを固定化できる（トピックマップだけで可能なわけではないが、上記のオントロジを応用した情報発見手法に加えてPSIが使える点が利点である）。

人名同士や、役職同士である特定のつながりがあるのかなどを分析することも可能となる。これにより、今までは漠として指摘されていた、造東大寺司写経所の人間関係などがわかる可能性がある。写経に従事する時期の差や作業状況なども、人名をセマンティックに分析することで可能となる部分もあるであろう。また、榮原永遠男氏が行った、月借錢解に関する、人間関係の分析なども、より精密に検討できる可能性を秘めている。

さらに、時間情報を入れておくことで、これらの関係性が時間の変遷とともに変わるのかなどの情報を得ることができる。上記の情報や研究をもととした古代国家の仏教編成などについて、中林隆之氏との共同研究を進めており、その研究の支援へとつながることも期待できる。

将来的な正倉院文書トピックマップでは、これ以外にも、多くの研究知識情報と正倉院文書の史

料そのものを有機的に結び付け、正倉院文書研究に関する知識情報の到達点を情報学的に提示すると同時に、次のステップへと役立てることを目指したい。

6. 正倉院文書トピックマップの課題と今後

上記のように、正倉院文書トピックマップはSOMODAの問題解決の技法を継承しつつ、特に写経所文書を中心として、研究のあり方などを可視化することを可能としている。

上記のような具体的な成果を上げると同時に、以下のような課題を持っている。⁽²¹⁾

<課題>

(1) 複雑な時間遷移の処理

正倉院文書は奈良時代という比較的短い期間の文書群ではあるが、時間的な変化はあり、かつ、その時間的な変化のなかでオントロジの構造が変わる。この変化をどのように処理していくかは今後の課題である。また、正倉院文書の記述内容そのものは奈良時代のものだが、その後の複雑な史料構造の変化まで含める場合には、江戸時代や明治時代まで考慮の対象とする必要がある。このような状況にどのように対応するのか、今後の課題といえよう。

(2) 適切なトピックの設定

第一段階では、人名・経典名といった史料に即した情報を入力しているが、トピックマップの本来の目的に沿うならば、より抽象的な研究知識情報を入力することが求められる。このトピックの設定が、正倉院文書研究にどこまで即しているのかを常に考慮しておく必要がある。

(3) 断簡IDの変化への対応

続々修については、調査が継続しており、今後断簡IDが、変更される可能性が高い。しかし、トピックマップでは「永続的識別子」を付すことが、より望ましいとされている。リンクされる先が可能な限り変わらないということも、研究資源としては重要であるため、今後どのような処理を行うのが適切なのか、検討をする必要がある。

次に正倉院文書トピックマップは、今後、以下のような発展を目指している。

<発展>

(1) オントロジの拡張

正倉院文書トピックマップは研究情報をより効果的に表現することを目指している。役職を構造化し、人物の関係を効果的に表現する、経典の関係を整理し、経典の構造を描く、研究語彙をオントロジ化するなどのより複雑なオントロジを応用することによって、より抽象的な研究知識情報の構築を目指す。これにより、過去の研究のあり方が可視化され、研究の発展につながったり、研究がわかりやすく可視化されることで、より効果的な正倉院文書研究のアウトリーチにつながることが期待できる。

(2) 他史料とのリンク

正倉院文書の同時代史料である『続日本紀』などとのリンクを構築することで、奈良時代の全体像をデジタルで記述することも可能となるであろう。

『続日本紀』については、現在、時空間情報を持つ記述を抽出し、HuTime⁽²²⁾への適用を試みるなどの動きを進めると同時に、TEI⁽²³⁾への適用実験を試みている。奈良時代の全体像を、より具体的にコンピュータ上で描くことが可能となるであろう。正倉院文書研究を奈良時代全体の研究と接続させることで、研究のステップをもう一段階先に進めることができると考えられる。

(3) 他の正倉院文書のリソースへのリンク

本稿執筆現在、Web 上における正倉院文書そのもののデジタルデータの提供はSOMODA⁽²⁴⁾のみである。しかし、いくつかの正倉院文書に関連する機関において、正倉院文書の高度情報化に関する検討が行われている。今後、これらの成果がリソースとして公開されるならば、それらのリソースとのリンクを作成することで、正倉院文書の関係する総体的なデジタルデータの提供が全体としては可能となると考えられる。

正倉院文書トピックマップは、正倉院文書の研究の方法論をデジタルで記述することを試みると同時に、SOMODAの発展形として位置づけている。SOMODAは、正倉院文書の復元過程を中心に記述したが、正倉院文書トピックマップは、そこからさらに研究資料の知識情報もあわせて記述することを目指した。

正倉院文書トピックマップは、情報技術やWebの研究動向にあわせた新しい人文情報研究の形を目指したものであるといえる。

7. 正倉院文書のデジタル化の意義—おわりにかえて—

以上、SOMODAの作成・課題から、正倉院文書トピックマップの事例について述べてきた。これらの事例が共通して目指したものと、変化したもの、そして現時点でもなしえていないことについて整理を行うことで、正倉院文書のデジタル化の意義へとつなげることにしたい。

SOMODAは、そもそも複雑な構造である正倉院文書を、コンピュータを用いて「復元」を行い、分散した情報を同時に閲覧することを目指したものであった。当初の狙いは、正倉院文書の復元・整理を行う際に用いられる「短冊」⁽²⁵⁾をコンピュータ上で表現することを目指した。その後、Webでのデータベースの公開が多くの基本となりつつある2000年代前半の状況の中、Webでのデータ公開へと方針を転換した。

次の正倉院文書トピックマップでは、基礎データの提供だけではなく、研究に関わる知識語彙の整理を試みている。これは、単純な検索ではなく、史料と語彙のリンクを複雑に組み合わせることで、より高度な検索結果を導き出すことを試みたものである。

この両者を通じて、一貫して進められていることは、「正倉院文書研究の方法論を情報技術によって記述する」試みである。正倉院文書研究は、1990年代以降、写経所文書を中心に非常に精緻な研究成果をあげている。しかし、これら精緻な研究成果を正倉院文書以外の研究者が参照しようとした際に、正倉院文書研究者と同じように複雑な手続きをトレースする必要があった。これは、「研究」として見ればやむを得ない部分があるのは事実ではあるが、正倉院文書はそのトレースがいわゆる日本古代史の他の研究とも異なる点で、より困難な状況を示している。このことは、正倉院文書研究のアウトリーチや、より学際的な研究の可能性を妨げる方向に働いているのではないかと考えている。

また、写経所文書研究は、ともすれば「(いささか言葉は強いが) 個別写経事業を追いかけるだけ」になってしまう可能性があり、外部との接続が一見困難な状況も起こってしまう。その際に、正倉院文書の研究手法をある程度情報技術によって記述し、それを Web で参照・再検証できるようになれば、上記のような課題を解決できるのではないかと考えている。SOMODA は正倉院文書の復原という史料操作の過程を、そして正倉院文書トピックマップは SOMODA に加えて研究そのもの(ある人物とある経典の関係性の整理など)を記述するという試みへと変化してきたものである。また、正倉院文書トピックマップは、それらの研究成果から、⁽²⁶⁾ tolog のような、トピックと関連に基づいた複雑な検索を可能とする式を用いることで、次の研究の「ヒント」を与えうるモデルを目指している。現在のようなある意味「シンプルなオントロジ」だけではなく、より複雑なオントロジを描くことで、上記のような「ヒント」を与えるモデルへと近づくのではないかと考えている。

しかし、このオントロジの記述は、ある研究の側面だけを記述したものになりがちである。ある僧侶と経典の関係を研究者 A が指摘したとして、それを後の研究 B で否定された場合、どちらの研究を記述すべきかなどを決定することは困難である。これは「通説的理解とはどのようなものか」といった問題にまで踏み込みかねない。「一つのある文献によって僧侶と経典の関係を説明できる」という説明が限界とならざるを得ないであろう。このように、研究を情報技術によって記述することは、研究成果の「固定化」でもあり、揺れ動きつづける研究のダイナミズムを記述するには、なお技術革新が必要であろうと考えられる。

一方、情報共有の観点から見た場合、正倉院文書は「恵まれた史料」であり、文書そのものだけではなく、それに関連する情報が非常に豊富である。また、多くの組織で、その目的に即した形でデジタル化が積極的に進められている。これらのデジタル化は、正倉院文書を研究・編纂もしくは閲覧といった目的に即した「記述」であり、その方法は組織ごとに少しずつ異なっている。それ自体は意味のあることではあるものの、現在の Linked Data 隆盛の時代に鑑みれば、これらの記述を総体的にまとめることで、「正倉院文書の世界」をデジタルデータで記述するという方向へいずればつながると考えられる。それは、正倉院文書研究が、今までと異なるアプローチを含みつつ次のステップへと進むという事であると考えられる。

正倉院文書のデジタル化は、日本古代史の他の史料に比べて進んでいると「はじめに」で述べた。それは日本古代史に関わるすべての研究の新たなモデルケースとしても位置づけられるようになるのではないだろうか。

謝辞：正倉院文書データベース作成の際には、正倉院文書データベース作成委員会のみなさま、宮内庁正倉院事務所・東京大学史料編纂所・国立歴史民俗博物館をはじめとする関係諸機関のご理解・ご協力をいただいた。また、トピックマップの技術的側面については、(株)ナレッジシナジーの内藤求氏より助言をいただいた。ここに記して感謝申し上げる。

付記：本研究は JSPS 科研費 25730199, 40182456 の助成を受けたものである。

註

(1)——正倉院文書データベース (SOMODA) に関しては、すでに多くのところで報告を行っており、本稿では概要を中心に述べるのみとするが、重複もあることをあらかじめお断りしておく。なお、SOMODA の説明については、後藤真「正倉院文書データベースと復元」(『アジア遊学』113号, 2008年) などが詳しい。

(2)——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』11-4, 2002年

(3)——正倉院文書データベース (SOMODA) <http://somoda.media.osaka-cu.ac.jp> 2012年12月10日閲覧

(4)——『正倉院文書目録』1~6 (2014年3月現在) 東京大学史料編纂所編, 東京大学出版会

(5)——『正倉院古文書影印集成』1~18 (2014年3月現在) 宮内庁正倉院事務所編, 八木書店

(6)——註(1)拙論文より

(7)——正倉院文書データベース作成委員会は以下のメンバーである。なお、所属は事業推進当時のものである。正倉院文書データベースの重要な特徴として、大学を中心としつつ、宮内庁正倉院事務所をはじめ、東京大学史料編纂所・国立歴史民俗博物館といった、正倉院文書に関わる諸機関と協業しつつ、データベース入力を進めていった点がある。これは、それぞれの資料をもった機関のみでは集約できないものを、大学の研究目的に即すことによって、データを効果的に集約できるという、一つの可能性を示したものであるといえる。

正倉院文書データベース作成委員会

●作成代表者

栄原永遠男 (大阪市立大学)

●作成分担者

柴山 守 (京都大学)

小倉慈司 (宮内庁書陵部)

杉本一樹 (宮内庁正倉院事務所)

山口英男 (東京大学史料編纂所)

山本幸男 (相愛大学)

吉岡眞之 (国立歴史民俗博物館)

(8)——断簡についての説明は栄原永遠男『正倉院文書入門』(角川学芸出版, 2011年) に詳しい。

(9)——小口雅史「日本古代史研究のためのオンライン・データベース」『日本歴史』740号, 2010年

(10)——それぞれの批判とその応答については、後藤真「デジタル技術による歴史情報の可視化の試み」(『日本古代の王権と社会』栄原永遠男編, 塙書房, 2010年)

にでも触れているので、そちらを参照していただきたい。

(11)——本文テキストの構造は、『大日本古文書』編年1~25巻のデータであり東京大学史料編纂所が作成した「奈良時代古文書データベース」のデータを高度利用している。これは、正倉院文書の現状でも原状でもなく、テキストを編年で並び替えたものである。

(12)——トム・ヒース, クリスチャンバイツァー著・武田英明ほか翻訳『Linked Data: Webをグローバルなデータ空間にする仕組み』(近代科学社, 2013年), なおLinked Dataは、ある種のルールを持って機械が読め、かつ意味のあるリンクによってWebの中で意味の世界と作り出すという一つの手法である。それにより、Web世界でそれぞれ別に持ち公開しているデータ同士が有機的に結びつきを持ち、全体で意味を持つようになる。人文科学研究に即して述べるなら、一つのデータベースの内部で、研究に関わるすべての情報を持つのではなく、研究に関するデータベースや資源同士が意味を持ったリンクでつながることで、大きな研究情報の提供モデルとなることを意識している。

(13)——ISO/IEC 13250-2, Information technology—Topc Maps—Part 2: Data model

ISO/IEC 13250-3, Information technology—Topc Maps—Part 3: XML syntax

ISO/IEC 13250-4, Information technology—Topc Maps—Part 4: Canonicalization

なお、日本語の関連書籍では内藤求『トピックマップ入門』(東京電機大学出版局, 2006年) が明解で理解しやすい。

(14)——正倉院文書トピックマップの技術的な詳細は、以下の論考を参照のこと。後藤真・内藤求「正倉院文書データベースへのトピックマップ応用による奈良時代知識情報構築の試み」(情報処理学会『人文科学とコンピュータシンポジウム論文集2012』2012年11月), 同「正倉院文書トピックマップへの研究知識情報充実による解析と情報発見の可能性」(情報処理学会『人文科学とコンピュータシンポジウム論文集2013』2013年12月)

(15)——紙については、所属(1: 正集・2: 続修・3: 続修後集・4: 続修別集・5: 塵芥・6: 続々修・9: 流出) + 帙(続々修のみ01~47, それ以外は00) + 巻(01~50, 2桁) + 紙番号(000~999, 000は題籤軸などの本紙以外の情報に利用) + 表裏(1: 現状表・2: 現状裏) の9桁のIDを付している。断簡は所属+帙+巻(紙と同

様のルール)+断簡1(『正倉院文書目録』で①と付されるもの、00~99)+断簡2(『正倉院文書目録』で(1)と付されるもの、00~99)+表裏(紙と同様のルール)の10桁のIDとなる。

(16)——「作者」は、漢訳仏典を示す表現としては適切ではないが、漢訳・疏類の著述などをすべてまとめて表現するために、あえてこのような表現を行っている。

(17)——大正新脩大藏経テキストデータベース(SAT)(<http://21dzk.l.u-tokyo.ac.jp/SAT/>) 2013年12月閲覧。以下同。

(18)——「華嚴宗布施法定文案」(『大日本古文書』11/557~568以下巻・ページを記したものはすべて『大日本古文書』)、「東大寺大修多衆并律衆布施法定文案」(11/568~589)、「俱舎宗写書布施勘定帳」(12/147~161)、「写書布施勘定帳」(12/42~99)など。上記史料が、現在実際入力したデータである。

(19)——科学研究費萌芽研究「SPSS等による日本古代史料の統計学的研究」研究代表者：榮原永遠男・課題番号14651069

(20)——科学研究費基盤研究(C)「経典目録よりみた古代国家の宗教編成策に関する多面的研究」研究代表者：

中林隆之・課題番号22520659

(21)——なお、これらの課題についての技術的側面などや詳細は註(14)の論文に詳しく述べてある。

(22)——京都大学地域研究統合情報センターが中心となって運営している「HGIS研究会」を中心として開発された時空間分析ツール。HuTimeのWebサイトは以下。(<http://www.hutime.jp/>)

(23)——後藤真「HuTime/Mapの日本史研究への応用の試み—続日本紀を題材に—」HGIS研究会編『HuTime/Mapを使った研究事例と将来展望』報告書、2012年

(24)——正倉院文書のほとんどを収録している『大日本古文書』(編年)は、東京大学史料編纂所より「奈良時代フルテキストデータベース」としてデータベースが公開されている。

(25)——「短冊」についての説明は榮原永遠男『正倉院文書入門』(角川学芸出版、2011年)に詳しい。

(26)——<http://www.ontopia.net/topicmaps/materials/tolog.html>

(27)——宮内庁正倉院事務所の「宝物検索」(<http://shosoin.kunaicho.go.jp/shosoinPublic/top.do>)より正集の画像情報が提供開始されている。

(花園大学文学部、人間文化研究機構連携研究員)

(2014年1月7日受付、2014年8月12日審査終了)



図2 SOMODAのトップページ



図3 SOMODAの断簡表示画面(復原時)
左右で表裏関係を表現しており、左側の面が復原された面である。

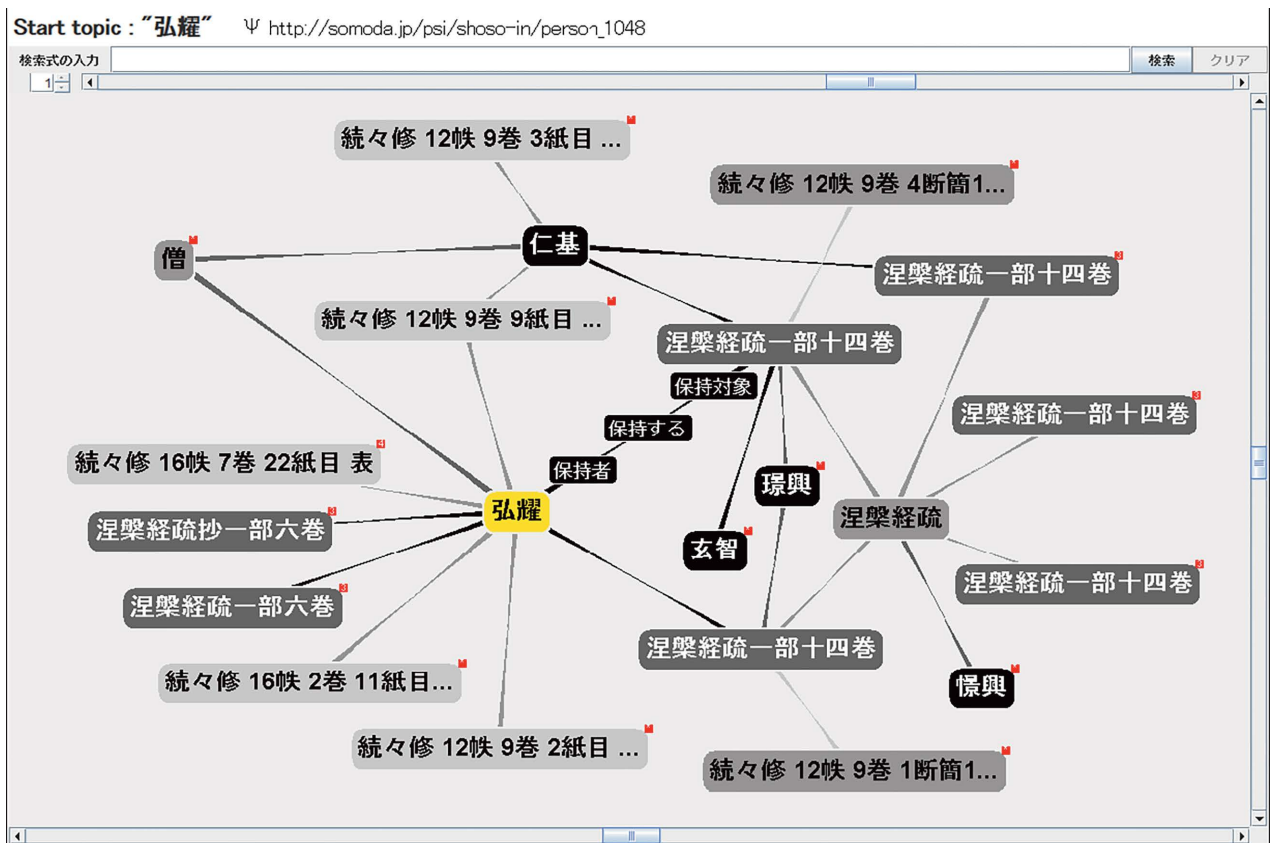


図5 僧名から經典名と個別經典・史料名の関連付けを Visualize したもの

以下のように定義されている（設定で変更可能）

- 黄色いノード（上記の例では弘耀）：焦点のあたっているノード，黒：人名ノード
- 濃いグレー：經典名ノード，グレー：断簡ノード，薄いグレー：紙ノード
- 線は関連型を示し，マウスオンで具体的な関連の内容を示している。上記の例では「弘耀」と「涅槃經疏一部十四巻」の関連の線上にマウスカーソルがあり，両者の関連を示している。