

情報資源共有化のための博物館資料 データベースのマッピングとその評価

Museum Materials for the Sharing of
Information Resources : Database Mapping and Its Evaluation

安達文夫・鈴木卓治・小島道裕・高橋一樹

はじめに

①情報資源共有化システムの概要

②マッピング方法

③システム構成

④実証システムによる評価

むすび

[論文要旨]

人文科学の分野において、様々なデータベースが作成され、多くがネットワークを介して公開されている。これらのデータベースをまとめて検索できるようにすることにより、個別のデータベースの所在やその操作方法を意識することなく検索が可能となる。

総合研究大学院大学の文化科学系の基盤機関と幾つかの大学とが共同し、各々の機関が有するデータベースを統合的に検索できるシステムの研究を進めている。この統合的な検索を実現するための一種の共通の窓として Dublin Core と呼ばれるメタデータを選択している。これに、国立歴史民俗博物館の「館蔵資料データベース」と「館蔵中世古文書データベース」のデータ項目をマッピングする方法について検討し、実証システムにより評価を行った。

ネットワーク上の資源の記述を本来の目的としたメタデータである Dublin Core に、資料の目録情報からなるデータ項目を一応の根拠をもって対応付けができる。しかし、資料の形状、状態、材質といった実際のものが持つ属性では、その意味の捉え方によって、対応付ける先にゆれが残る。また、ユーザインタフェース上、エレメントの名称を直接的に検索語の入力欄に表記したのでは、実際に対応付けられているデータの内容とかけ離れる場合が生ずる。エレメントの定義や名称を適正化する必要があることが、実証実験により確認された。検索結果の表示方法の評価から、表示するデータ項目や名称の表示方法の検討が必要であり、統合検索と個別検索の役割の整理が重要な課題であることが示された。また、多数のデータベースを対象とすることから、検索結果の表示と絞り込みに関するユーザインタフェースの面で新たな考慮が求められることが明らかとなった。

はじめに

人文科学の分野において、様々なデータベースが作成され、多くがネットワークを介して公開されている。これらのデータベースの検索によって、研究論文や調査報告書の存在、研究対象とする資料の所在を知ることができるとともに、資料を画像あるいは全文テキストとして収録したデータベースにより、その内容までを知ることができる。現状では各々のデータベースを個別に検索することになるが、複数のデータベースをまとめて検索できるようにすることにより、個別のデータベースの所在やその操作方法を意識することなく検索が可能となる。

分散して存在する情報資源を、相互に利用できるようにする試みは、図書館の分野では早くから始まり、この横断検索が実現している。博物館資料や記録史料の相互利用をめざした研究が行われるとともに [1-3]、複数の機関が所蔵する歴史資料を共通化して検索するシステムが構築されている [4]。そして、より広く人文科学系の諸機関のデータを共有化する研究が進められている [5]。

人文科学系のデータベースは、対象が多様で、それぞれの観点からデータ項目が与えられている。これを統合的に検索するには、いわば共通的な窓を通して、一致するデータを探すというアプローチをとることになる。この場合、有効な統合検索を実現するには、共通的な窓をどのように構成するかということと、この共通的な窓に、個々のデータベースのデータをどのように対応付けるかが重要な課題となる。

総合研究大学院大学の共同研究として、その文化科学系の基盤機関と幾つかの大学とが共同し、各々の機関が有するデータベースを統合的に検索できるシステムの研究を進めてきた [6]。そして、共通的なデータの窓として Dublin Core [7] と呼ばれるメタデータを選択し、システムを構築して実証実験を行ってきた [8]。国立歴史民俗博物館からは、所蔵資料の目録を収録している「館蔵資料データベース」と「館蔵中世古文書データベース」をその対象として選び検討を進めてきた。

共通的な窓への対応付けを、各 DB の個々のデータに戻って行うことは理想ではあるが、新たにデータベースを作成することと同等の手間がかかり、現実的ではない。そこで、多少の意味の食い違いや曖昧が生ずることは犠牲にして、対応付けのルールを定め、自動的に変換（マッピング）を行うことが実際的である。博物館資料に関して民族標本資料対象としたマッピングの検討が行われている [2]。歴史資料に関するデータベースについても、マッピングのルールをどのように定めるかが課題となる。

本論文では、二つのデータベースを Dublin Core へのマッピングを検討する中から、多様な研究情報資源を統合的に検索することを目的に選択したメタデータに、歴史資料の目録情報をマッピングする上での考え方を明らかにするとともに、実証システムを通しての評価をもとに、マッピングの適合性と統合検索システムの課題を明らかにする。

①……………情報資源共有化システムの概要

1.1 構成概要

情報資源共有化システムは、国文学研究資料館、国際日本文化研究センター、国立民族学博物館、東京大学史料編纂所、京都大学東南アジア研究所、大阪市立大学、並びに国立歴史民俗博物館が有しているデータベースを対象としている。具体的なデータベースは文献 [8] に示されているが、これらは、収録する分野が広範なだけでなく、対象が文献や、各地に所在する研究資源、あるいは所蔵する資料・図書と多様であり、また所蔵資料も実物の資料の他、マイクロフィルム等と幅が広い。そしてその内容も、目録情報、所在情報、資料のフルテキスト、画像と多岐にわたる。このように多様なデータベースを共通のメタデータを媒体として共有化を図り、統合的な検索を実現する。メタデータとしては、図書、記録史料、博物館資料それぞれに幾つかが提案され利用されているが、対象に依存しないという意味で比較的汎用的な Dublin Core を選択している。

情報資源共有化システムの基本的な構成を図 1 に示す。上記のデータベースを横断的に検索するため、Z39.50 と呼ばれるプロトコルを適用している。Z39.50 とインターネットのプロトコルを交換する機能を持つ Z39.50 ゲートウェイ (Z39.50GW) を、幾つかの研究機関が設置している。利用者はこのゲートウェイにアクセスすることにより、Web のブラウザにより横断検索サービスを利用することができる。

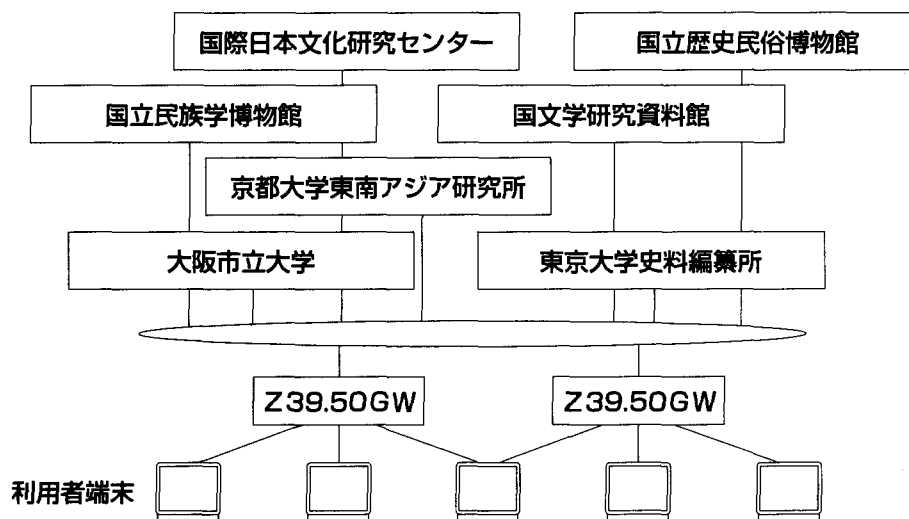


図 1 資源共有化システムの基本構成

1.2 Dublin Core の適用方針

ここでは、共同研究のグループとしての方針を記す。Dublin Core は、表1に示す15項目の基本エレメントから構成される。Dublin Core の規約では、この基本エレメントに Qualifier を付けることにより拡張できる。しかし、この情報資源共有化の研究に当たっては、基本要素だけの Dublin Core で、多様な人文科学系のデータベースをどこまで統合的に検索できるかを評価することを一つの目的として、Qualifier を用いないことを前提とした。

表1 Dublin Core のエレメント

DC element	意味
Title	リソースの名称
Creator	リソースの制作者（機関）
Subject	リソースの内容が対象とする題材
Description	リソースの内容の説明
Publisher	リソースを利用可能とした機関や者
Contributor	リソースの制作に寄与した機関や者
Date	リソース自身の出来事に関する日付
Type	リソースの内容の性質やジャンル
Format	リソースの物理的・デジタル形式の説明
Identifier	リソースを明確に識別する記述
Source	現在のリソースの源となるリソースへの参照
Language	リソースの内容を記述する言語
Relation	関連するリソースへの参照
Coverage	リソースの内容に関わる時間的、空間的な範囲
Rights	リソースの権利に関する情報

時間に関わるエレメントとして、Date と Coverage がある。人文科学系のデータベースを利用するに当たっては、時期あるいはその幅を指定した検索が必要である。このような検索が可能となるように、西暦表記による年・月・日のデータを該当するエレメントに配することとした。このため、元のデータベースで和暦や時代で表記されているものは、西暦表記への何らかの変換を行うことが必要となる。

この情報資源共有化システムにより目的とする対象が検索されたときに、元のデータベースの画面を指定により表示可能とすることを方針とした。これにより、元のデータベースにリンクされている画像の閲覧等が可能となる。これを実現するため、元のデータベースのデータへのリンク情報をいずれかのエレメントに割り付けることが必要となる。

②……………マッピング方法

国立歴史民俗博物館では、大別して、所蔵する資料の目録情報を収録した館蔵資料データベース、研究分野毎の文献や共同研究等で調査、収集した成果をまとめた集成的研究データベース、主に記録類を対象とした全文データベースを公開している。一方、情報資源共有化システムで対象とする他の機関のデータベースは、文献資料を対象とするものが多い。統合検索の効果を評価するには、

複数の機関のデータベースでヒットすることが必要である。このことから、館蔵中世古文書データベースを選択した。そして、もの資料のデータも収録している館蔵資料データベースをもう一つの対象とした。

以下では、まず情報資源共有化の対象とした二つのデータベースの特徴について記し、これらのデータ項目をマッピングする上での基本的考え方を述べ、具体的なマッピング法を記して検討を行う。なお、館蔵資料データベースを「館蔵資料DB」、館蔵中世古文書データベースを「中世古文書DB」と記す。

2.1 館蔵資料DBと中世古文書DBの概要

今回対象とした館蔵資料DBと中世古文書DBの二つのデータベースは、性格が異なっており、それぞれの意味について説明する。

(1) 館蔵資料DB

館蔵資料DBは、文字通り国立歴史民俗博物館が所蔵する約20万点の資料すべてを一元的に検索できるようにしたものであり、資料管理を行うことを目的とした目録を基にしている。そのためデータ項目は、国立歴史民俗博物館に収蔵されている文献・考古・民俗など多様な資料の、いわば最大公約数的なものに限定している。この一覧を表2に示す。資料名称・コレクション名という名称関係の項目、「重文」などの文化財指定区分、数量・法量・材質・実物模造の別、といった物体としての側面を示す項目、製作年代・使用地という、資料としての内容的な属性に関わる項目、そして管理のための資料番号（便宜上、A（考古）・F（民俗）・H（歴史）の三つの記号を付し、あとは登録順に番号を付けている）、および備考がある。以上の区分に入らないデータや、研究上の所見などは、注意事項などとともに、「備考」に適宜記入される。

表2 館蔵資料データベースのデータ項目

データ項目	説明
資料名称	資料の名称
コレクション名	資料が含まれるコレクションの名称
区分	国宝、重要文化財、重要美術品、重要有形民俗文化財、都道府県指定の区分
数量	収蔵数
法量	資料の縦、横、高さ、重量
材質	資料の材質
実物模造	実物、模造の区分
尺度	模造の場合の製作比率
製作年代	西暦区分、西暦年、世紀、時代、元号年による表記
使用地	所在地、出土地
資料番号	管理番号
備考	該当の資料に関する注意事項、調査所見等

(2) 中世古文書DB

館蔵資料DBでは、個々の資料ジャンルによっては不可欠な情報であっても、共通的に設けた項目に当てはまらない場合は「備考」に記入される。たとえば、年代は「年」までしか記入する項目

がないため、文献資料の月日は備考への記入となり、また古文書には不可欠な差出人や宛先といった情報も同様である。このため、これを基に検索するには不都合が生ずる。このような問題を補完する機能とともに、それぞれの資料（群）の属性に即した、より詳細なデータを提供することを目的として、個別のデータベースを作成・公開している。中世古文書DBもそのひとつに位置づけられる。

中世古文書DBは、国立歴史民俗博物館が所蔵する日本中世（平安時代末期～戦国時代）の古文書を、学術的な目的での検索にも耐えるように、中世古文書という資料の特性に合わせて設計したデータベースであり、中世古文書の目録を作る際に使用する調書の内容を前提にしている。そのデータ項目とその意味を表3に示す。

まず管理番号については、番号は一点ずつ収集されたバラの文書にはそのまま付くが、ひとかたまりの文書群であればそれに対して付けられ、さらにそれは複数の文書から成る巻物や冊子によって構成されている場合があるため、この場合特定の一点の古文書を同定するためには、資料番号—卷子・冊子番号—文書番号という三段階の番号を必要とする。

表3 館蔵中世古文書データベースのデータ項目

データ項目	説明
資料番号	当該文書を含む所収文書群（下記参照）の整理番号
卷子・冊子番号	当該文書を含む所収文書群のなかでの卷子または冊子の整理番号
文書番号	当該文書の整理番号。卷子または冊子に含まれている場合は、卷子または冊子ごとに文書一点ずつ順番に番号を付与
文書名	古文書学にもとづく文書の名称。つづけて括弧内に当該文書の正文・案文・写の区別、現状の装頓が作成当時から後世のものかの区別を表示
欠簡・断簡	文書の前部や後部などに欠損がある場合は欠簡。文書の欠損が甚だしく伝存する形状が断片的である場合は断簡
日付1	当該文書に記載されている和暦年・月・日
日付2	当該文書に記載されている和暦年を西暦年に直し月・日とつづけて数字列で表示
西暦	当該文書に記載されている和暦年を西暦年で表示
形式	古文書学にもとづく当該文書の料紙の形状や使用法の区分を表示
時代	当該文書の作成された時代。原本ならば文書に記載される日付と時代は一致するが、案文や写しの場合はそれが作成された時期の時代を表示
法量縦	文書の縦の長さ（最大数値） 単位はcm
法量横	文書の横の長さ（最大数値） 単位はcm
紙数	文書に使用された紙の数
差出	文書に記載される差出者
充所	文書に記載される宛先の人名もしくは機関・団体名
端裏書	文書の右端の裏側に記入されたメモランダム
本文書出	文書に記載された文章の冒頭部分
本文書止	文書に記載された文章の末尾に書く文言
備考	当該文書に記載された地名などの関連情報
所収文書群	当該文書を含む家伝文書群やコレクションの内部でのまとまり。おもにコレクションにおいて旧蔵者を単位とする。
区分	当該文書を含む家伝文書群やコレクションの名称

文書名は様式や内容、また原本か写しか、といった情報から、学術的な基準で付けられた名称である。それに欠損があるかないか（欠簡・断簡）であるかの情報が付随する。

日付・年代関係では、四つの項目がある。

まず「日付1」は、資料の記載通りに、和暦で年月日を記したものである。「応仁二年二月廿一日」のようになる。

しかしこれでは年代順の並べ替えができないため、便宜的に11けたの数字で表したのが「日付2」である。先の例なら、14680002021となる。上4桁の西暦の後2桁は、年代が推定である場合に「99」を入れ、また閏月は、下3桁目に「5」を入れて対応している。

「西暦」は、4桁の西暦年のみを入れてあり、これによって何年から何年までという絞り込みを容易に行うことができる。

「時代」は、物としての古文書が作成された年代であり、「室町時代中期」といった、幅のある表現になっている。先の三つの項目が古文書の記載された内容の年代であり、古文書に記載されている年月日ないし推測した年月日を記しているのに対して、この「時代」は記載された年月日とは必ずしも一致しない。つまり、古文書が原本であれば、通常は記載された年月日が作成された年月日だが、原本ではなく後に写されたものである場合、作成すなわち写された年代は記載された日付とは異なることになる。記載されている日付は室町時代でも、「時代」は江戸時代である、という場合が実際にはしばしばあるのである。

「形式」は、文書として使われている紙の形状や使用のしかたにもとづく古文書学上の分類をしめしているが、その分類のあり方は単に料紙のかたちにも着目して独自に行われるものではなく、データベース上で差出（発信者）・充所（受信者）や本文書出・本文書止などに表現される、古文書の様式や内容と密接に関わっている。

分量も館蔵資料DBであれば、表具や巻物の大きさが重要だが、中世古文書DBでは、当然一点一点の古文書本紙の寸法が記載される。また、複数の紙を継いだりである場合があるため、「紙数」も記される。

「差出」と「充所」は、文書史料を定式的に記入できる内容として最も重要な情報である（学術用語としては、「文書」とは発信者と受信者がある文字史料のことである）。

これ以外の内容に関わる情報としては、畳まれた文書を開かないでも分かるように当時の人が裏に書いたメモ＝「端裏書」と、本文の「書出」・「書止」の部分項目としている。書き出し部分には文書の主旨が書かれることが多く、また書き止めは、様式や関係が表れやすい部分で、文書の性格を理解する上で重要である。また、備考欄にも、文書に出てくる荘園名など、内容の理解に関わる情報が記される。

古文書は一点だけが伝来することもあるが、関係文書と共にまとまって伝わる場合が多く、またコレクターの手によってまとめ直される場合もある。この点についての情報として、「所収文書群」があり、ある一点の文書がどのような文書のかたまりの一部であるのかを知ることができる。「区分」の項目は、さらに大きな括りであり、「越前島津家文書」といった家わけの名称や、「田中稔氏旧蔵典籍古文書」といった、コレクションの名称である。田中本（コレクション）の中に「山科家文書」の一群がある、といった関係が現実にあるため、このように二種類の項目になっている。

2.2 マッピングの基本的考え方

博物館資料とそのデータベースの関係を図2に示す。実物の資料に加え、文献資料の写し・写本や、展示や研究のために製作された複製がある。これらの写真・マイクロフィルムも重要な研究資源である。この情報を書き留めた目録があり、これを電子化したものがデータベースとして存在している。これを共通的な項目にマッピングするものである。

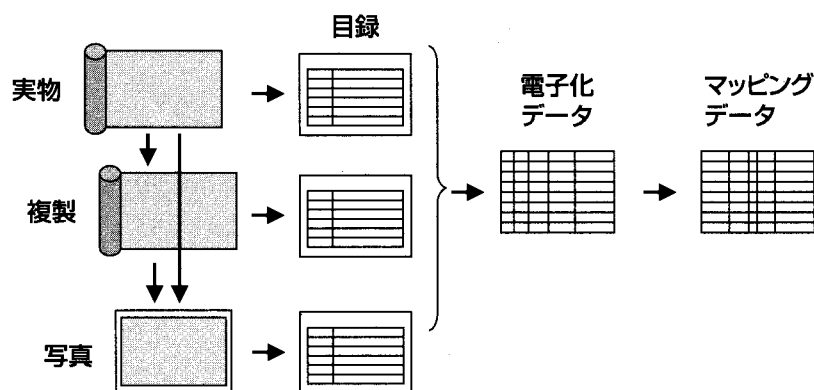


図2 博物館資料とデータベースの関係

Dublin Core は、ネットワーク上の資源を相互流通のために記述することを元来の目的としている。このネットワーク上の資源ということ厳密に適用すると、図2の中で、目録を電子化したデータがその資源ということになる。このとき、例えば Creator はそのデータの作成者、Date はその作成日付となって、資源共有化サービスの利用者である研究者に意味を持たない情報となる。このため、館蔵資料 DB では資料そのもの、中世古文書 DB では文書そのものをリソースと捉え、Dublin Core のドキュメントの記述を解釈することとした。

横断検索の Protokol として適用した Z39.50 の機能により、検索の対象とする項目（以下、検索用）と検索の結果として返す項目（以下、返戻用）を分けることができる。そこで、Dublin Core のエレメントへマッピングする際に、検索用と返戻用の使い分けを考慮することとした。

表2と3に示した館蔵資料 DB と中世古文書 DB のデータ項目は、Dublin Core のいずれかのエレメントに対応付けを行うことを基本とした。これが適正かは、実証システムにより評価を行う。

2.3 Dublin Core へのマッピング

Dublin Core へのマッピングを検討するにあたっては、Dublin Core Metadata Initiative (DCMI) のドキュメント [7] を参照するとともに、Consortium for the Computer Interchange of Museum Information (CIMI) のガイドライン [9] を参考にした。以下では、まずこの二つのドキュメントを基にしたマッピングを記した後、これにより残ったデータ項目のマッピングについて検討を加える。次に、元のデータベースではデータ項目として有していない情報の付与について述べる。以下で、出典を示さず「～が示されている。」とあるのは、文献 [7] に依っている。最終的なマッピング結

果を、表4と表5に示す。

表4 館蔵資料DBのマッピング結果

DC element	検索用	返戻用
Title	資料名称 資料名称 カナ	資料名称 資料名称 カナ
Creator		
Subject	コレクション名 コレクション名 カナ ‘最上位を除く上位の資料番号の資料名称’	“国立歴史民俗博物館所蔵資料” コレクション名 コレクション名 カナ ‘最上位を除く上位の資料番号の資料名称’
Description	区分 数量 材質 備考	“区分：”区分 “数量：”数量 “材質：”材質 “備考：”備考
Publisher		
Contributor		
Date		
Type	実物・模造	実物・模造
Format	法量 尺度	法量 尺度
Identifier	資料番号	資料番号
Source		
Language		
Relation	コレクション名	コレクション名 ‘リンク情報’
Coverage	製作年代 ‘YYYY’ 使用地	製作年代 ‘YYYY’ 使用地
Rights		“国立歴史民俗博物館所蔵”

“ ”:固定値, ‘ ’:生成値

表5 館蔵中世古文書DBのマッピング結果

DC element	検索用	返戻用
Title	文書名	文書名
Creator		
Subject	区分 所収文書群	“国立歴史民俗博物館所蔵”区分 所収文書群
Description	差出 充所 端裏書 本文書出 本文書止 備考	“差出：”差出 “充所：”充所 “端裏書：”端裏書 “本文書出：”本文書出 “本文書止：”本文書止 “備考：”備考
Publisher		
Contributor		
Date	時代 ‘YYYY YYYY’	時代 ‘YYYY YYYY’
Type		
Format	形式 欠簡・断簡 法量縦, 法量横 紙数	形式 欠簡・断簡 “縦：”法量縦“横：”法量横 “紙数：”紙数
Identifier	区分“-”資料番号“-”卷子・冊子番号 “-”文書番号	区分“-”資料番号“-”卷子・冊子番号 “-”文書番号
Source		
Language		“ja”
Relation	所収文書群 区分	所収文書群 区分 ‘リンク情報’
Coverage	日付1 日付2 西暦	日付1 日付2 西暦
Rights		“国立歴史民俗博物館所蔵”

“ ”:固定値, ‘ ’:生成値

2.3.1 ドキュメントに基づくマッピング

(1) Title

館蔵資料 DB では資料名称を，中世古文書 DB では文書名を対応させる。

(2) Creator

Creator は著者，制作者を記載するエレメントである。館蔵資料 DB では，制作者が判明している資料では，備考のデータ項目に記述されている。人手により対応付けを行うのであれば，このエレメントに入力できるが，機械的なマッピングを前提としているため，このエレメントは使用しない。中世古文書 DB においても制作者に関するデータ項目はなく，対応付けを行わない。

(3) Subject

このエレメントには，資源の内容に関わるトピックを，キーワードやフレーズ，あるいは分類を用いて記述するよう示されている。館蔵資料 DB は，図3に示されるように，コレクション毎に，枝番で資料番号が与えられ付与され，それぞれに資料名称が付与されたツリー構造になっている。ある資料の上位の資料名称は，キーワードとして有用であることから，これを Subject に対応付ける。中世古文書 DB の区分と所収文書群も同様にキーワードして有用であることから Subject に対応付ける。

なお，館蔵資料 DB において，最上位階層の資料名称とコレクション名は同一であるから，重複することのないよう，一方だけを対応づける。

(4) Description

資源の内容に関する記述や説明を記載することが示されている。中世古文書 DB の差出，充所，本文書出，本文書止，端裏書をここに対応付ける。二つの DB の備考には，資料の内容に関わる記述があることから，これを Description に対応付ける。

また，CIMI のガイドラインでは，材料や技法の記載に Description を使うことが示されており，館蔵資料 DB の材料を Description に対応付ける。

(5) Publisher

館蔵資料 DB，中世古文書 DB とも該当するデータ項目はなく，対応付けを行わない。

(6) Contributor

館蔵資料 DB，中世古文書 DB とも該当するデータ項目はない。

(7) Date

Coverage の項目に記す。

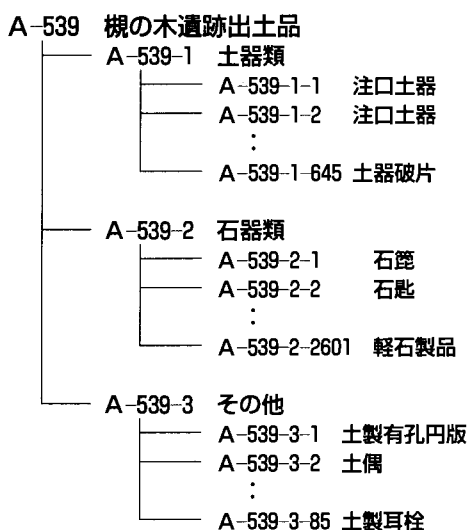
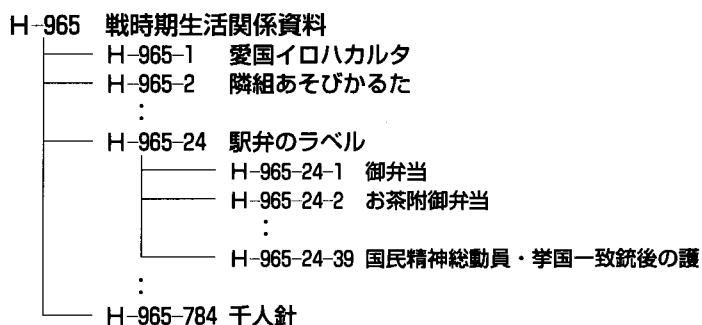
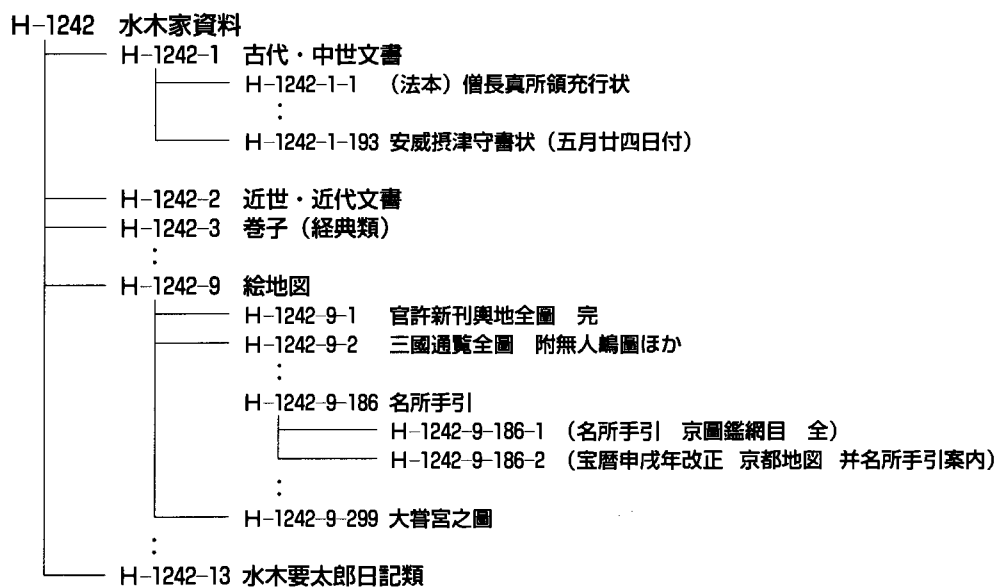


図3 館蔵資料群の構造

(8) **Type**

Type は、資源の内容に関する性質や分野を記述するもので、定められた一覧の中から選択するとされている。その例として、Image, Dataset, Text 等が挙げられている [10]。この意味で両データベースとも Type に対応付ける項目はない。一方、CIMI のガイドラインでは、異なった分野の博物館のコレクションの検索の助けとなるよう、実物と代用品 (surrogate) をその一覧に加えるべきとしている。これに準じて、館蔵資料 DB の実物・模造を Type に対応付けを行う。

(9) **Format**

Format は資源の表示や再生に必要な機器を特定するために物理的な形式やメディアの種類を記載することを目的としたエレメントである。資料の単なる大きさの記述に当てはまるかは明確ではない。CIMI のガイドラインでは資料の大きさ (dimension) を記載することが示され、幾つかの例示がある。これに倣い、館蔵資料 DB の法量 (寸法)、尺度と中世古文書 DB の法量 (寸法) を Format に対応付ける。

(10) **Identifier**

館蔵資料 DB では資料番号を対応させる。中世古文書 DB では、一つだけで文書を特定する識別子を与えるデータ項目はない。館蔵中世古文書資料群は、先に記したとおり、区分、所収文書群、所収文書群が分かれている場合は卷子・冊子、ならびに文書を指定することによって文書を特定できる。そこで、所収文書群、卷子・冊子と文書には識別子にあたるデータ項目を用い、識別子が付与されていない区分はその表記をそのまま使用して、区分+資料番号+卷子・冊子番号+文書番号を Identifier として対応付ける。

(11) **Source**

このエレメントは、資源が派生した元となるものの参照先を記すものである。複製や写本の場合、その元となる資料を Source に記述すべきであるが、機械的にマッピングを行うため、このエレメントは使用しない。

(12) **Language**

資源を記述する言語を示すものであり、両 DB とも対応付けるデータ項目はない。

(13) **Relation**

関連する資源の参照となる情報を記述することが示されている。館蔵資料 DB ではコレクション名を、中世古文書 DB では区分と所収文書群を対抗付ける。館蔵資料 DB の最上位を除く上位階層の資料名称は、固有の名称でなく普通名詞による記述もあるため、Subject のように上位階層の資料名称の全てを対応付けることは行わない。

(14) Coverage

Coverage は資源が表現している内容に関する時間や空間の情報を、Date には資源が制作されたあるいは利用可能になった日付を記載することが示されている。館蔵資料 DB の使用地は資料の内容に関わる場所ではないが、資料と結びついた場所とも言えるので Coverage に対応付ける。同 DB の製作年代は、文字通りであれば、Date に対応する。しかし、複製品において、それが制作された年代ではなく、原品が制作された時代が記載されていることもあり、その資料の持つ内容に関わる時間情報となっている。原品が制作された年代も資料を説明する情報の一つであることから、一律に扱うため、Coverage に対応付けている。

中世古文書 DB の日付は、原本であれば Date に対応するが、写しはここへの対応付けは適さない。一律に Coverage に対応付けを行う。文書が作成された時期を記載した時代を Date に対応付ける。

館蔵資料 DB の実物・模造のデータ項目の内容により、対応付け先を変えたり、中世古文書 DB の日付を文書名より正文であるかを判定して Date に代入することも考えられるが、このような厳密な対応付けを行う必要があるかは、今後の課題である。

(15) Right

両 DB とも対応付けるデータ項目はない。

2.3.2 未対応づけの項目の検討

以上、Dublin Core と CIMI のドキュメントに沿って対応付けを行った。対応付けが終わっていないデータ項目は

館蔵資料 区分, 数量, 法量 (重量)

中世古文書 形式, 欠簡・断簡, 紙数

である。これらは、資料の形状, 状態, ならびに物理量であり、ネットワーク上の資源では有しない属性である。CIMI のガイドラインに基づいて対応付けを行った材料, 実物・模造, 法量 (寸法), 尺度も同様な性質を持つ。

この対応づけが残った項目について、何らかの根拠を持ってマッピングを行わなければならない。資料の物理量を表す項目である館蔵資料 DB の法量 (重量) と中世古文書 DB の形式, 紙数は、Dimension の意味を拡大解釈して Format に対応付けた。それ以外の館蔵資料 DB の区分, 数量と中世古文書 DB の欠簡・断簡は記述の一種と見なして Description に対応付けを行った。ここで、館蔵資料 DB の数量は個別の資料の属性ではなく資料群に対する属性として意味を持つことから Description に対応付けている。

2.3.3 自動生成・固定割付けの情報

情報資源共有化システムでは、横断検索した結果を表示する画面から、元のデータベースの対応する資料 (または文書) の検索結果画面を表示できるようにする。元のデータベース画面へのリンク情報は、資源を参照する有力な情報であることから、対応付けるエレメントは Relation が適切

と考えられる。但し、実装では、Identifier に対応付けた資料番号あるいは区分～文書番号の文字からリンクを貼っている。

データベースを個別に検索するときは、利用者はその結果が何のデータベースからのものか当然了解している。ところが、横断検索の場合は、そのユーザインタフェースによっては、どのデータベースからの結果か分からない状況が生じる。このため、どのような研究資料群を記述するデータか明らかとなるよう Subject にその名称である“国立歴史民俗博物館所蔵資料”と“国立歴史民俗博物館所蔵”+区分名を固定的に割り当てることとした。

Language には資源を記述している言語を統制語で記述することが示されている。中世古文書 DB では、対象とする資源が全て日本語で記述されていることから“ja”を固定的に割り当てた。館蔵資料 DB では文字が記載されていない資料があり、一律の割り当てができないため、割り当てを行わない

Rights は諸権利を記載する。ここでは二つのデータベースとも、このエレメントに“国立歴史民俗博物館所蔵”と固定的に割り当てた。

以上の固定的に割り当てた項目は、全件ヒットを避けるため、検索用には与えず、返戻用にだけ与えている。

なお、CIMI のガイドラインでは、博物館が資料を収蔵し公開する行為を一種の出版と捉えて、Publisher に博物館名を割り当てることが示されている。しかし、収蔵する資料や文書を研究資源として記述する立場から、この記法は取り入れていない。

2.3.4 マッピングの考察

以上のように、二つのデータベースのデータ項目を一応の根拠を持って対応付けることができる。しかしながら、資料の大きさや状態、材料を記載するエレメントに関しては、ガイドラインとしてあっても、定義として明確化されていないことから、考え方を少し変えれば、別のエレメントへの対応付けもあり得ることとなる。例えば館蔵資料 DB の材質は Description に対応付けたが、Qualifier に関するドキュメント [11] では、Format の Qualifier として Medium を定義し、材料 (material) を記すとしている。これに準ずれば、材質は Format にマッピングすることになる。また、中世古文書 DB の形式は、袋綴や折紙など文書の料紙の形状を表すことから Format に対応付けた。しかし、これが物理的な形状より、文書の役割や機能を表すことの意味が重要であると考えれば、マッピング先として Description が適する。

このように、捉え方によって対応付ける先が変わり得ることは、資料の所蔵機関によって（あるいは人によって）同一にマッピングされない—ゆれを持つ—ことを意味する。そして、利用者から見た場合、どこにマッピングされているか事前に知ることができない、あるいは利用者の捉え方によって、想定するエレメントにデータ項目が対応付いていないという問題を生み出す。これは、検索結果の表示を人が見て利用する上では、大きな支障とはならない。しかし、検索結果を機械的に利用する場合や、ゆれを持つデータ項目が検索のキーとして使用され場合には解決を必要とする。

2.4 年代・日付の表記と変換

時間に関するデータとして、館蔵資料 DB では製作年代、中世古文書 DB では日付と時代がある。複数のデータベースを統一的に検索することを目的とした西暦年への変換法について記す。

年月日の表記方法については、Dublin Core で YYYY-MM-DD とハイフンで結んだ数字で表記することが推奨されている。年だけであれば、YYYY である。幅を持つ時期の表記、何年“頃”の表記、紀元前の表記については、規約はないが、CIMI のガイドラインに記載された事例をもとに、それぞれ YYYY/YYYY, ca.YYYY, YYYYBC と表記することにした。この表記方法については、今後定められる規約に基づいて改める必要がある。

(1) 館蔵資料 DB の製作年代

このデータ項目には、次の5種類の表記が用意されており、必要に応じて一つ以上を記載するようになっている。

- (a) 西暦区分 (AD, BC)
- (b) 西暦年
- (c) 世紀1, 世紀区分1, 世紀2, 世紀区分2
- (d) 時代1, 時代2
- (e) 元号, 元号年

元号は、元号年なしに記述できる。時代は表6の中から選択する。時代は時代1だけの表記と、年代2を併記して年代の幅を表す表記がある。世紀1, 2の表記も同様である。世紀区分はA~Dの記号で1/4世紀毎を示し、Xは“頃”を示す。このような方法による次のような記載がある。

例1 : AD, 1324, 世紀 : 14-A, 鎌倉時代, 元号 : 元亨-04 年

例2 : AD, 江戸時代, 元号 : 寛永

例3 : AD, 世紀 : 17-BX~17-CX, 江戸時代

例4 : AD, 鎌倉時代

例5 : AD

このような多様な年代表記に対して、西暦年への変換則を定める必要がある。ここでは以下の原則で変換則を構成した。西暦年があれば、これを使用する。西暦年と元号年がなく、元号がある場合は、その幅を表現するよう変換する。これらが無い場合は、世紀か時代を使用する。両者の記載がある場合は、目途として50年より短い年代、具体的には安土桃山時代、明治時代、大正、昭和1, 昭和2, 平成が時代1だけで記載されているときは時代を優先し、これ以外は世紀および世紀区分の記載を優先する。弥生時代は西暦区分で分けることができるので、世紀の記載がない場合は、西暦区分を利用する。世紀、時代とも記載がない場合は、西暦区分から生成して表記する。以上による変換則を図4に示す。世紀と時代の両者の記載がある場合、その組合せにより、細かな期間を生成できる。例えば14世紀、鎌倉時代とあれば、1301/1333を生成できるが、ここでは取り入れていない。この必要性があるかは、今後の課題である。

以上により変換する際に、時代の開始と終了は表6のテーブルより生成する。ここで、平安時代

```

if 西暦年≠NULL then 西暦年を代入 西暦区分が BC なら BC を付加
else if 元号年=NULL and 元号≠NULL then 元号の対応表の値を代入
else if 時代2=NULL
    and 時代1=安土桃山時代
    or 明治時代
    or 大正
    or 昭和1
    or 昭和2
    or 平成 then 時代の対応表の値を代入
else if 世紀1-1≠NULL then 世紀1, 2, 世紀区分1, 2により値を生成して代入
else if 時代2=NULL
    and j 時代1=弥生時代 then YYYYBC/1BC   YYYY は時代の対応表による
    and 西暦区分=BC
else if 時代2=NULL
    and 時代1=弥生時代 then 1/YYYY   YYYY は時代の対応表による
    and 西暦区分=AD
else if 時代1≠NULL then 時代1と時代2により, 時代対応表をもとに値を生成
    して代入
else if 西暦区分≠NULL then 西暦区分に応じて/1 BC or 1/を代入
else 何も代入しない
    
```

図4 製作年代から西暦年への変換

表6 館蔵資料DB 製作年代“時代”一覧

項目	開始	終了
旧石器時代		10000BC
縄文時代	14000BC	400BC
弥生時代	950BC	250
古墳(飛鳥・白鳳)時代	250	700
奈良時代	701	794
平安時代	794	1185
鎌倉時代	1185	1333
室町時代	1333	1569
安土桃山時代	1569	1600
江戸時代	1600	1868
明治時代	1868	1912
大正	1912	1926
昭和1 (戦前)	1926	1945
昭和2 (戦後)	1945	1989
平成	1989	

表7 館蔵資料DB 製作年代“元号”変換表(部分)

元号	開始	終了	元号	開始	終了
元徳	1329	1332	応安	1368	1375
元弘	1331	1334	建徳	1370	1372
正慶	1332	1334	文中	1372	1375
建武	1334	1338	永和	1375	1379
延元	1336	1340	天授	1375	1381
暦応	1338	1342	康暦	1379	1381
興国	1340	1346	永徳	1381	1384
康永	1342	1345	弘和	1381	1384
貞和	1345	1350	至徳	1384	1387
正平	1346	1370	元中	1384	1392
観応	1350	1352	嘉慶	1387	1389
文和	1352	1356	康応	1389	1390
延文	1356	1361	明德	1390	1394
康安	1361	1362	応永	1394	1428
貞治	1362	1368	應永	1394	1428

～江戸時代の開始は、館蔵資料データベースの「検索の手引き」にある年によった。元号の名称を時代としているものは、検索もれが起こらないよう、終了と次の時代の開始とが同一年となるようにテーブルを構成した。時代の開始・終了の利用者の理解に幅があると考えられるものは、やはり検索もれが起こらないよう、時期が重複するテーブルとした。テーブルの書き換えにより研究の進展に対応できる。

元号の変換テーブルの一部を表7に示す。同一年に二つの元号を持つ南北朝時代も、このテーブルで変換できるが、同一の名称を持つ“建武”は北朝と南朝の識別が付かないので、期間の長い方の値をテーブルに与えている。また、同表にある“応永”と“應永”のような元号の表記のゆれは、本来、元のデータベースで統制すべきであるが、テーブルに加えて西暦年の生成を行った。

(2) 中世古文書 DB の日付, 時代

日付は, 検索用に用意されている“西暦”を使用する。時代は, 表 8 に示すような内容が入力されている。その種類は50程度であることから, 同表に示すように時代の開始と終了の年を示すテーブルを作製した。これを参照して, YYYY/YYYY を生成する。

表 8 館蔵中世古文書 DB “時代” 変換表(部分)

内容	開始	終了
平安時代	794	1184
平安時代～南北朝時代	794	1391
平安時代中期	901	1003
平安時代後期	1004	1086
平安時代院政期	1087	1184
平安院政期～南北朝時代	1087	1391
平安院政期～室町時代後期	1087	1569
平安末～室町時代	1087	1569
鎌倉時代前期	1185	1221
鎌倉時代	1185	1332
鎌倉時代～南北朝時代	1185	1391
鎌倉時代前期～南北朝時代	1185	1391

③……………システム構成

国立歴史民俗博物館における実証実験システムの構成を図 5 に示す。システムは Z39.50サーバとデータベースサーバとから構成される。Z39.50サーバは PC サーバを, データベースサーバはデータベースエンジンを搭載した UNIX サーバを用いた。

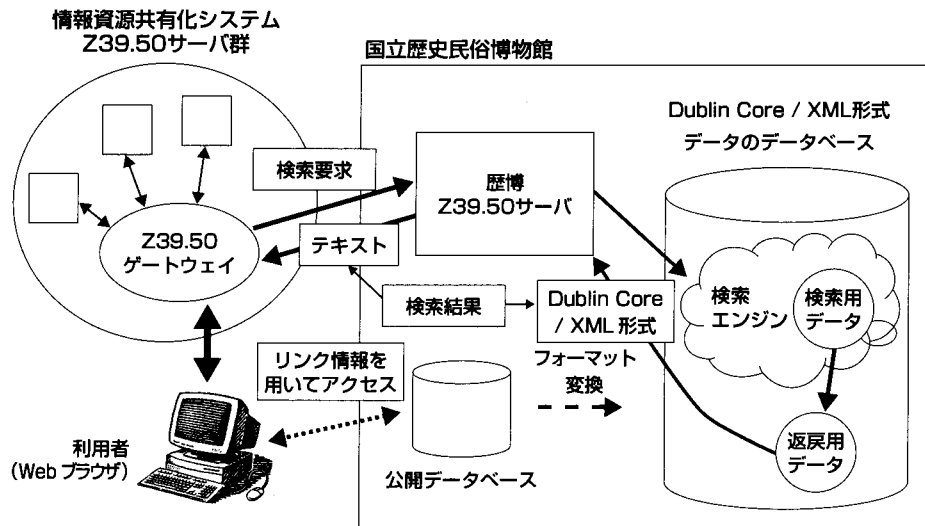


図 5 情報資源共有化システムの構成

国立歴史民俗博物館に設置された Z39.50サーバは、特定の Z39.50ゲートウェイとの通信のみ許諾されるよう、ネットワーク側でセキュリティ保護を行なっている。利用者は、情報資源共有化システムに参画している幾つかの機関で用意されたゲートウェイを利用して国立歴史民俗博物館のデータベースにアクセスすることになる。

Z39.50サーバは、検索のリクエストを Z39.50ゲートウェイから受け取ると、データベースを検索し、検索結果を Z39.50ゲートウェイに返す、という作業を受け持っている。実証実験システム用のデータベースとして、公開データベース上のデータを、Dublin Core にマッピングし、XML 書式でタグ付けした構造化データ (Dublin Core/XML 形式と称す) を作成し、これをデータベースに管理させる。このとき、検索用のデータと返戻用のデータの2種類を作成し登録しておく。データベースは検索条件を受け取ると、検索エンジンを用いて検索用のデータを探索し、検索条件に合致したレコードについて、返戻用のデータを検索結果として Z39.50サーバに返す。返戻用のデータには、公開データベースの当該データに直接アクセスするためのリンク情報が埋め込まれている。Z39.50サーバは、この検索結果を、直接表示可能なテキスト形式に直して、呼び出し元である Z39.50ゲートウェイに返す。その際、リンク情報は「http://～」の形の文字列に変換されるが、これがブラウザ上でクリック可能なリンクとして機能するかどうかは、Z39.50ゲートウェイの実現に依存する。

公開データベースのデータから、Dublin Core/XML 形式のデータを得る過程について、中世古文書 DB を例に説明する。まず、公開データベースから抽出した CSV 形式のデータ (図 6(a)) に、仮のタグ付けを行なう。この時点で、JIS 外文字など規格外の文字の利用についてチェックし、必要ならばデータを修正し、さらに、日付や年代情報など、検索に必要で、かつ自動生成が可能な情報を加えたデータ (図 6(b)) を作成する。これをもとに、XSLT により、Dublin Core/XML 形式のデータを、検索用、返戻用それぞれについて生成する。この際、表示用データに公開データベースへのリンク情報を書き込む。図 6(c)に検索用の、図 6(d)に返戻用の Dublin Core/XML 形式データを示す。検索用のデータが SGML 書式で与えられているのは、データベースエンジンの仕様による。各図中のアンダーラインの部分は、各段階の処理によって追加・変更が行なわれた部分を強調して示したものである。

このように生成されたデータをもとに検索が行われる。利用者の端末から見た画面の遷移を図 7 に示す。また、検索の結果として表示される詳細表示画面の例を館蔵資料 DB と中世古文書 DB それぞれについて、図 8 と図 9 の(a)に示す。これよりリンクが張られ最終的に表示される元のデータベースの画面を各々の(b)に示している。

1008,384,1,1,僧本性敷地寄進状 (改装) ,,康永四年八月三日,13450008003,竖紙,南北朝時代,30,42,1,僧本性 (花押) (目下), (金光寺) (文中), (前欠) のちのけん/そうのきしんしやう,奉寄進金光寺 (七条高倉) 敷地事,仍寄進状如件,,七条道場金光寺旧蔵古文書第一巻,田中本古文書,1345

(a) CSV 形式

図 6 検索用データの生成法

```
<?xml version="1.0" encoding="Shift_JIS"?>
<record-list>
  <BAN>1008</BAN>
  <SIRBAN>384</SIRBAN>
  <KANBAN>1</KANBAN>
  <MONBAN>1</MONBAN>
  <MONMEI>僧本性敷地寄進状 (改装) </MONMEI>
  <HIDUK1>康永四年八月三日</HIDUK1>
  <HIDUK2>13450008003</HIDUK2>
  <HIDUK2>1345</HIDUK2>
  <KEISIK>  堅紙</KEISIK>
  <JIDAI>南北朝時代</JIDAI>
  <JIDAI>1333/1391</JIDAI>
  <TATE>30</TATE>
  <YOKO>42</YOKO>
  <SISU>1</SISU>
  <SASDAS>僧本性 (花押) (日下) </SASDAS>
  <ATEDOK> (金光寺) (文中) </ATEDOK>
  <HASIUR> (前欠) のちのけん / そうのきしんしやう</HASIUR>
  <KAKDAS>奉寄進金光寺 (七条高倉) 敷地事</KAKDAS>
  <KAKTOM>仍寄進状如件</KAKTOM>
  <MONGUN>七条道場金光寺旧蔵古文書第一巻</MONGUN>
  <KBN>田中本古文書</KBN>
  <SEIREK>1345</SEIREK>
</record-list>
```

(b) original/XML 形式

```
<!DOCTYPE record-list SYSTEM "dc-tyuusei">
<record-list>
  <dc-record>
    <title>僧本性敷地寄進状 (改装) </title>
    <subject>田中本古文書</subject>
    <subject>七条道場金光寺旧蔵古文書第一巻</subject>
    <description>僧本性 (花押) (日下) </description>
    <description> (金光寺) (文中) </description>
    <description> (前欠) のちのけん / そうのきしんしやう</description>
    <description>奉寄進金光寺 (七条高倉) 敷地事</description>
    <description>仍寄進状如件</description>
    <date>南北朝時代</date>
    <date>1333/1391</date>
    <title>僧本性敷地寄進状 (改装) </title>
    <format>  堅紙</format>
    <format>30</format>
    <format>42</format>
    <format>1</format>
    <identifier>384-1-1</identifier>
    <relation>七条道場金光寺旧蔵古文書第一巻</relation>
    <relation>田中本古文書</relation>
    <coverage>康永四年八月三日</coverage>
    <coverage>13450008003</coverage>
    <coverage>1345</coverage>
  </dc-record>
</record-list>
```

(c) Dublin Core/SGML 形式 (検索用)

図 6 検索用データの生成法(つづき)

```
<?xml version="1.0" encoding="Shift_JIS"?>
<record-list>
  <dc-record>
    <title>僧本性敷地寄進状 (改装) </title>
    <subject>国立歴史民俗博物館所蔵田中本古文書</subject>
    <subject>七条道場金光寺旧蔵古文書第一巻</subject>
    <description>差出：僧本性 (花押) (日下) </description>
    <description>充所：(金光寺) (文中) </description>
    <description>端裏書：(前欠) のちのけん/そうのきしんしやう</description>
    <description>本文書出：奉寄進金光寺 (七条高倉) 敷地事</description>
    <description>本文書止：仍寄進状如件</description>
    <date>南北朝時代</date>
    <date>1333/1391</date>
    <title>僧本性敷地寄進状 (改装) </title>
    <format>縦紙</format>
    <format>縦：30 横：42</format>
    <format>紙数：1</format>
    <identifier>384-1-1</identifier>
    <language>ja</language>
    <relation>七条道場金光寺旧蔵古文書第一巻</relation>
    <relation>田中本古文書</relation>
    <identifier><![CDATA[<IMG SRC="http://www.rekihaku.ac.jp/icon_web.jpg"
ALIGN="absbottom"><A HREF="http://base1.nijl.ac.jp/~metadata/cgi-bin/dblink2.cgi?DBNAME=tyuusei&KEYWORD1=%93c%92%86%96%7B%8C%C3%95%B6%8F%91&KEYWORD2='384'&KEYWORD3=1&KEYWORD4=1" TARGET="websearch">田中
本古文書-384-1-1</A>]]></identifier>
    <coverage>康永四年八月三日</coverage>
    <coverage>13450008003</coverage>
    <coverage>1345</coverage>
    <rights>国立歴史民俗博物館所蔵</rights>
  </dc-record>
</record-list>
```

(d) Dublin Core/XML 形式 (返戻用)
図6 検索用データの生成法 (つづき)

④……………実証システムによる評価

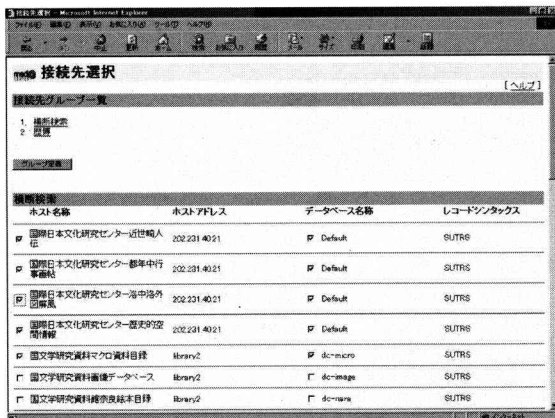
この章では、著者の中で、資源共有化システムの利用者の立場となる二人の評価を記し、次に課題を考察する。

4.1 利用者としての評価

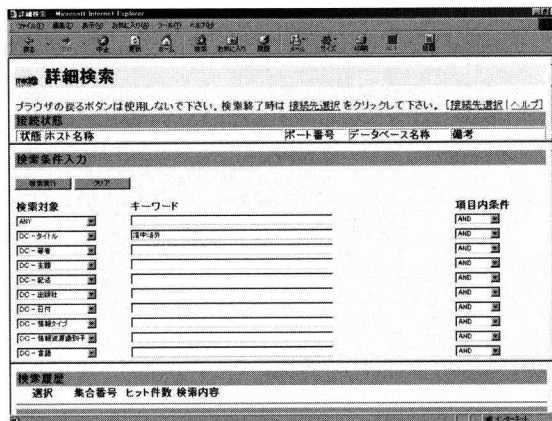
(1) 意見1

実現したシステムをどう評価するかは、おそらく利用する個人によって大きく異なると思われるが、以下それを前提に、個人的評価として述べる。

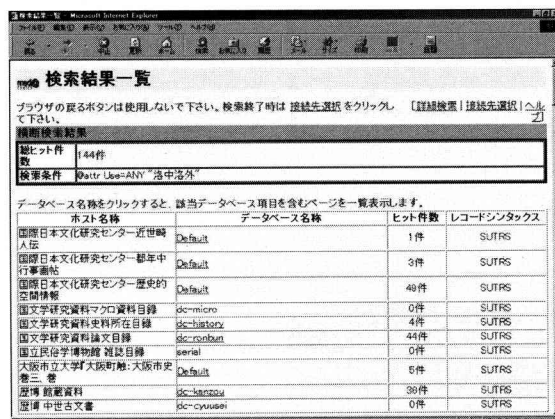
まず、効果としては、言うまでもなく、各機関のデータベースを横断的に検索できることによつて、どのようなリソースがあるかを即座に知ることができることである。検索した項目に対して、自分が知らなかった意外なものがヒットすることで、思いがけない資料や文献の所在を発見するこ



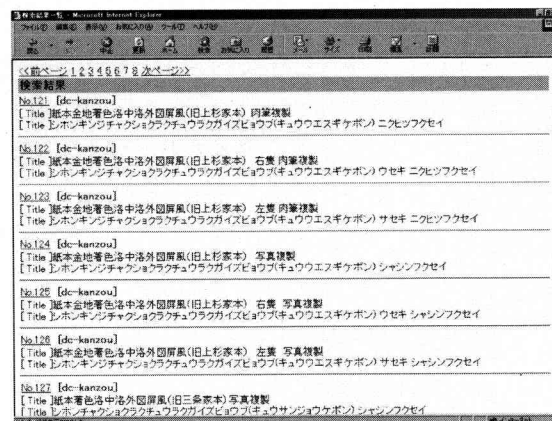
(a) DB 選択



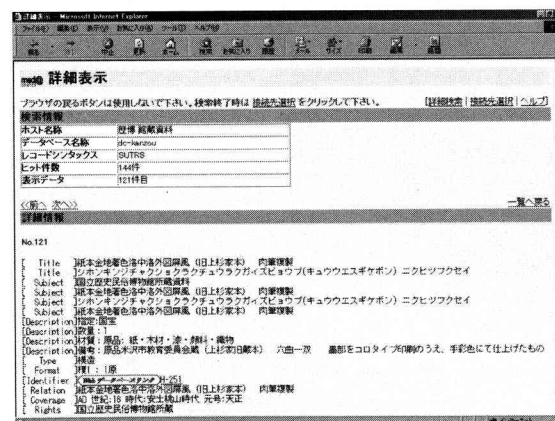
(b) 検索語入力



(c) DB 毎ヒット件数表示

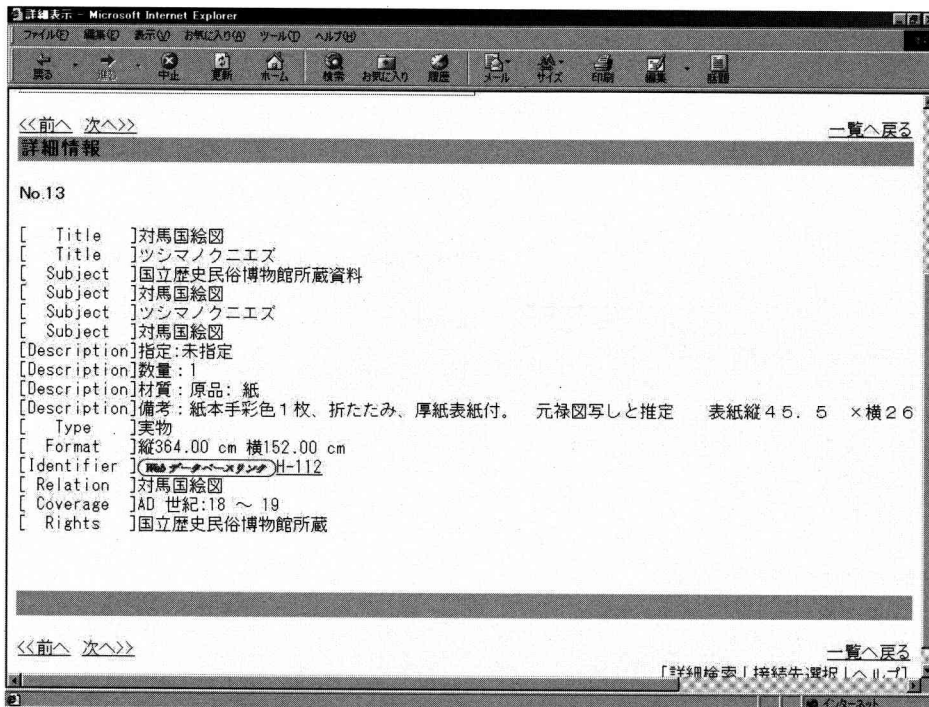


(d) 一覧表示

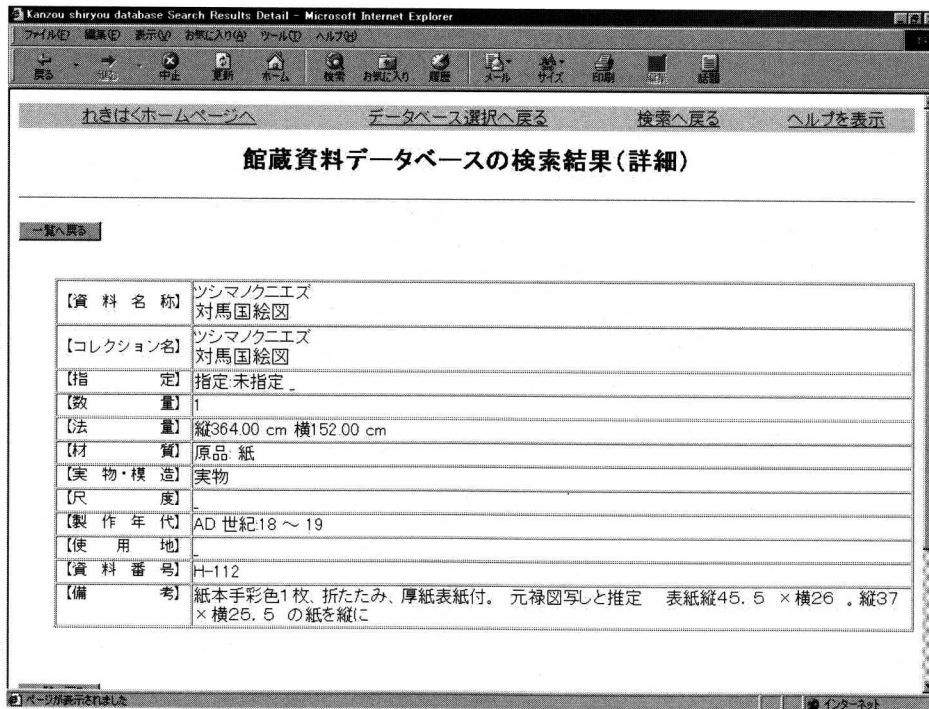


(e) 詳細表示

図7 検索の画面遷移

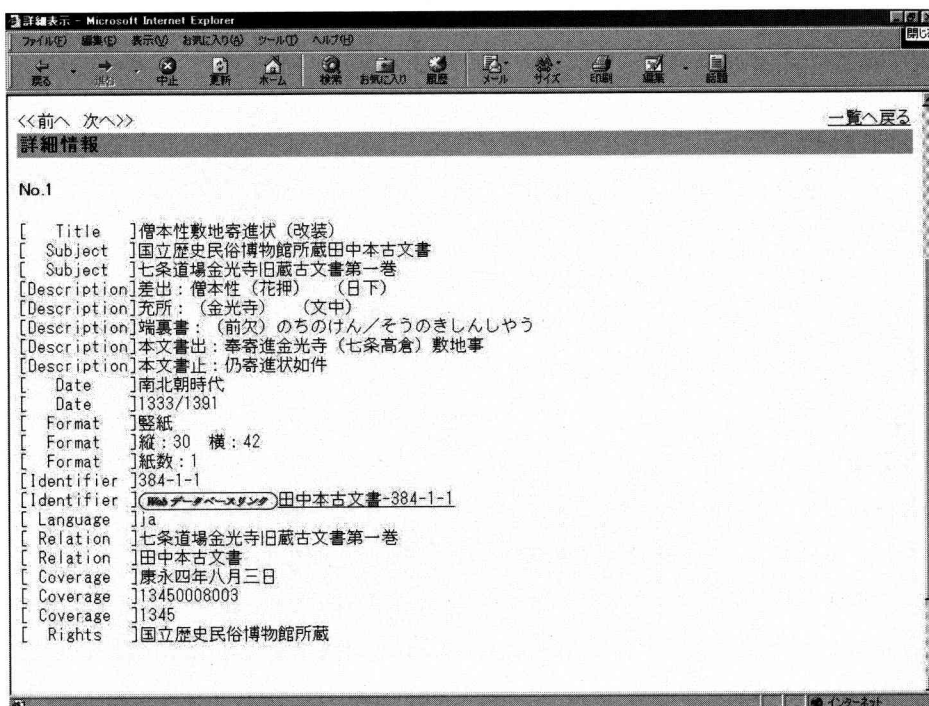


(a) 資源共有化システムの表示画面

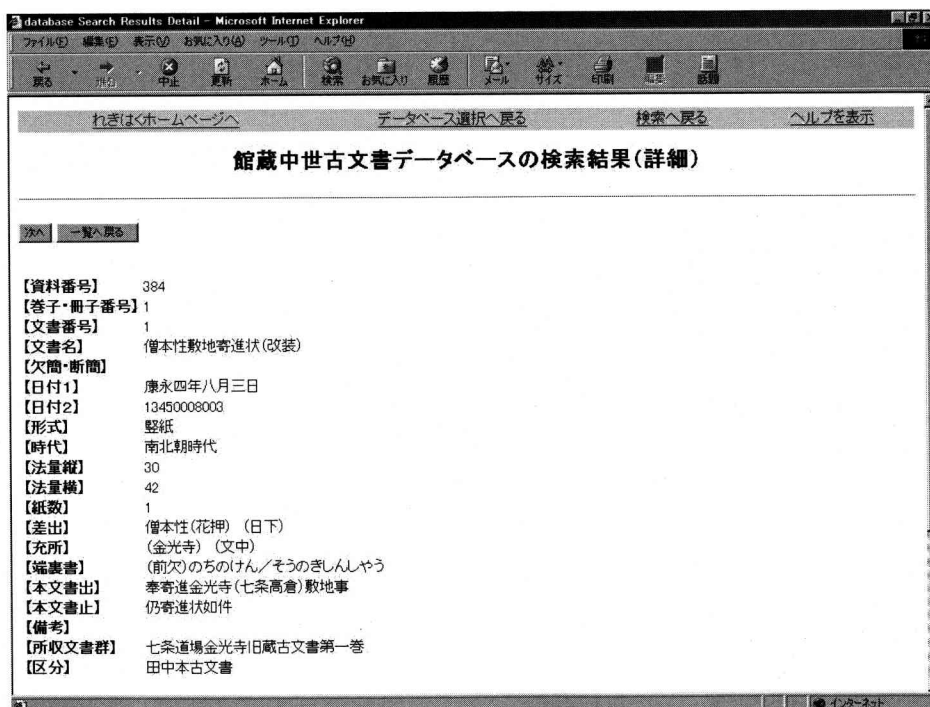


(b) 元のデータベースの表示画面

図8 館蔵資料DBの詳細表示画面例



(a) 資源共有化システムの表示画面



(b) 元のデータベースの表示画面

図9 館蔵中世古文書 DB の詳細表示画面例

とができる効果は大きい。

しかし、自分が探そうとしているものが本当に検索されるかという点では、やや不安が残る。以下に記すように、個別のデータベースを使用する場合に比べて、検索が必ずしも思うようにはできないからである。

特に「ダブリンコア」による検索項目については、残念ながら違和感がきわめて強く、項目の中身をイメージすることができないため、あまり役に立たない。実質的に使えるのは、「ANY」か「タイトル」くらいであろう。あとは図書の場合の「著者」と「出版年」くらいだが、図書の検索だけなら別に専用のデータベースがあるから、この横断検索を用いる意味はあまりない。筆者にとって特に利用したいのは時代による絞り込みだが、「日付」の欄はどのように入力するのがわからないし、おそらく各データベースによってまちまちだから、検索する側としては事実上使うことができない。結局のところ、任意のキーワードを「ANY」か「タイトル」に入れてみる、というのが実質的な利用方法ということになり、それ以上の検索は難しかった。

次に、検索結果の表示についてだが、これはそれぞれのデータベースに固有の項目がそのまま表示された方がわかりやすい。それぞれのデータベースは、言うまでもなく対象とする資料等について、その属性を分析して項目をつくっているため、結局のところ、その思想と秩序を理解しながら使うほかはないし、それは実際にその資料等の一つの有効な項目分類になっているはずである。それをダブリンコアの違和感の強い項目に無理に置き換えることは、その秩序をわざわざ崩す行為に他ならず、利用者にとってはよりわかりにくくされただけの結果に終わっていると一言をえなない。今回表示されるデータベースの中で、ダブリンコアに置き換えられたデータベースはわかりにくく、もとの項目がそのまま出てくるDBの方がわかりやすい。結局のところ、ダブリンコアの項目は、表に出す必然性はほとんどなく、かえって利用者に不便を強いるだけのように思われる。

全体として言えば、横断検索には、個々のデータベースを使う場合とは異なった利用の仕方が必要となるため、利用者は困惑することが多い。そして、結局のところ、個々のデータベースについての理解がなければ有効な利用は難しい。

横断検索を実質的に効果があるものとするためには、このような点をユーザインターフェースの面で強力にサポートしていくことが不可欠と言えよう。すなわち、どうすれば何を知ることができるのかを明示し、また、個々のデータベースがどのような目的で、どのような資料を集めているのかを容易に知りうるようにすることが必要であると思われる。

(2) 意見2

今回の共同研究は、複数の研究機関が公開する30以上の多様なデータベースの横断検索を可能とする実験であった。その意味では、各データベースの接続から検索まで比較的軽快に行うことができ、大きな成果をあげたものと考えられる。ただしその一方で、今後の課題とすべき点も少なくなく、おもにユーザの立場からインターフェースの改善を中心に気づいた点を述べておきたい。

まず第1に、検索の方式に関して。検索のために入力する語句が例示されておらず、専門的な研究者以外の一般ユーザには使いづらい仕様となっている。また、いわゆる絞り込み検索の方法が不明確であること、とくに横断検索につきものの膨大なヒット件数からムダなデータをいかに除外し

ていくかの方策が不十分である。そもそも、接続するデータベースの選択が検索対象となるデータベースの選択にそのまま直結していることが、使いづらさの大きな原因になっている。

データベースの接続と検索の分離に関わっては、現状のデータベース接続の選択方式が30以上のチェックボックスへの個別入力となっている点の改善を前提として、すべてのデータベースの横断検索の選択も排除すべきではないが、それ以前に必要なのは、データベースの内容にもとづくグルーピング（たとえば一次資料と二次資料、目録データベースとフルテキストデータベースなどの区別）を複数設定して、より効率的にユーザーが求めるデータの取得がはかれるよう支援すべきである。この点と深く関わるが、データベースのグルーピングを提示するにあたっては、各データベースのディレクトリを充実させることが重要となろう。

第2に検索結果の表示に関して、図7に示されるとおり、この横断検索はあらかじめ接続したデータベースを対象にフリーワードを含む検索を行うと、その検索語句を含む①データベースの一覧表示と②個別のデータを一覧表示（ただし表示件数は20件）し、最終的には各機関のデータベースにリンクするようになっている。

しかし、横断検索の結果表示②において、データベース名称が表示されるのみでは不十分であり、ホスト名称も併記すべきである。また、前述のように、この横断検索は接続するデータベースの選択が検索対象となるデータベースの選択と同一であり、しかも接続＝検索したデータベースの順番と表示件数に規定されて検索結果が表示されるため、すぐに検索結果を知りたいデータベースを選択しても、表示された同じページに他のデータベースの検索結果が混じるという不都合が生じる。検索結果の表示件数が変更可能なか否かもわかりにくい。これらの点の改善を行うとともに、検索結果のソート機能の充実が必要である。

なお、より具体的な問題点としては、データベース間における字体の対応関係にまで議論が立ち入れなかったことは反省すべきである。また、横断検索の結果から各機関のデータベースのローカルデータにリンクすることが可能となることから、歴博のデータベースに関する個別の課題となるが、詳細データ表示から画像表示の画面遷移を別のウィンドウで立ち上がるようにして、詳細データと画像とを並列的に参照できることが望ましい。

4.2 課題の考察

(1) 検索語入力での課題

検索語は、基本的に入力画面のDublin Coreの各エレメントに対応する入力欄から投入する。このとき、意見1および意見2に示すように、何を入力してよいか分からないという状況が生ずる。これは、文字の種類や形式についてどのように入力するかという問題と、どこに何を入力するかという問題に分かれる。そして、操作説明を完備すれば解決する問題と、操作説明や入力サンプルの例示では本質的には解決しない問題—検索語の入力欄の名称が利用者にとって適切かという問題—に整理できる。

操作方法を読まなくともある程度は利用できるようにすることは重要である。その意味で入力欄の名称は適切に与える必要がある。検索語の入力欄のラベル名称がエレメントの英語表記あるいはこれをそのまま訳した日本語表記では、利用者がどこに何を入力したらよいか分からないというこ

とが起こる。例えば、Coverageに時間の情報が入力されていることは、Dublin Coreをよく理解していないと分からず、Dateに入力してヒットしないことが起こりうる。時間、場所の情報が入力されていることが明確となるよう例えば“記載範囲（時期、地域）”としても、coverageの意味をそのまま表しただけでは、館蔵DBの製作年代や中世古文書DBの日付が入っていることは、名称から想像することは難しい。同様に、Subjectを主題と表記したのでは、キーワードに相当するデータが入っているとは理解しづらい。また、Creatorを著者としたのでは、制作者として入力する際に違和感を与える。

また、先に記したように、Dublin Coreの本来の目的がネットワーク上のリソースの記述であることに起因して、実物の資料の形状、状態、材料に関わる属性のマッピング先に、その捉え方によって、ゆれが生じる問題がある。

この利用者がどの入力欄を使って検索すればよいか不明となる問題を解決するには、各エレメントの意味を見直し、必要に応じて拡張すると共に、利用者にはエレメントに入るデータの内容に見合った名称で示すことが必要となる。

一方、仮にデータベースを構成する側で最適なメタデータを設計し、エレメントに入るデータの内容に相応しいラベル名称を付与できたとしても、利用者のラベル名称に対する意味の解釈に広がり存在し、多様なデータベースを共通的な項目に対応付けた項目のどれを選択するかゆれが残ることは避けられない。エレメントの定義と名称を適正化した上で、さらに複数のエレメントを必要に応じて同時に検索するようなシステムによる対応が必要となる。

(2) 検索結果表示の課題

意見1の検索結果をマッピング先のエレメント名で表示することが分かりづらいという問題は、Dublin Coreというより、多様な視点から構成されているデータベースを共通的な項目で表示することから生ずる問題である。

今回の二つのデータベースの結果表示において、図9に示すように、数字だけではその意味が伝わらない法量のようなデータ項目や、差出、充所、本文書出・書止のように内容だけでは区別がつかないものは、元のデータベースの検索結果表示と同一のデータ項目名を一種の見出しとしてデータの前に付けている。これを全てのデータ項目に適用し、さらに共通項目としているエレメントの名称を表示しなければ、上記の問題を回避できる。

しかし、幅広い研究資源を対象とした統合検索において、共通の項目名を示さず、元のデータベースのデータ項目名だけを表示することは、その意味を理解していない状況で表示されることもあり得て、利用状況によっては不適切なインターフェースとなる。検索語入力 of の課題と共通するが、項目名の表記の適正化による解決が必要である。

一方、元のデータベースの画面へのリンクを基本とするシステムであれば、統合検索の結果として全てのデータ項目を表示することは必要なく、探そうとする研究資源の候補の判断に必要な範囲が表示されればよいとも考えられる。統合検索と個別検索の役割の整理を図った上で、適切な検索結果の表示手法を確立してゆくことが課題となる。

(3) 全項目マッピングの評価

今回のマッピングにおいては、全てのデータ項目を Dublin Core の少なくとも一つのエレメントに対応付けることを基本とした。このため、数字だけのデータがマッピングされるエレメントがある。Format には、館蔵資料 DB の法量と尺度、中世古文書 DB の法量と紙数が入力されている。このため、例えば紙数の検索条件としてある数字を入れても、これと文字列が同一のデータがヒットし、いわゆるゴミが多くなりほとんど意味を持たない。意味を付随させることができない数字だけのマッピングは検索用としては行わず、これによる詳細な検索は元のデータベースに任せるといった統合検索の役割の整理がこの観点からも必要である。

なお、共通な検索を目的に変換した西暦年を表示しているが、元のデータ項目の表示と内容が重複することから、返戻用には割付ける必要はないと認められる。

(4) 時間に関する検索

時間情報を基に検索する要求は意見に示したとおり高い。今回の検討では、西暦年表記に基く共通的な検索が可能となるよう、その変換を行っている。しかし、記法は統一されたものではなく、今後標準化が必要である。また、幅を有する期間の検索について、指定の区間に収まるものと、指定の区間を含むものといった検索を実現する必要がある。

今回は対象としなかったが、考古学や民俗学の分野のデータベースでは、多くが場所の情報を有し、これに基づいた検索が求められる。時間情報と同様に、その表記法の標準化が必要であるとともに、区域の指定法を確立することが課題である。

検索対象として重要な、時間に関わるデータは数字で表記され、場所に関わるデータも少なくとも緯度経度は数字で表記される。両者が同一のエレメントに併存しても、或いは他の数字データと一緒に識別が付くよう、共通に検索する対象としての時間と場所のデータに対して、Qualifier の導入が必要である。

(5) ユーザインタフェースの課題

意見 2 のデータベースの選択とグルーピング並びに絞り込みの問題は、多数のデータベースを対象とすることから生じ、これまでの単独のデータベースの検索では起こらなかった問題である。従来の検索語ならびに検索項目の選択に加えてデータベースの選択という操作を旨く組み合わせることのできるインタフェースを実現してゆく必要がある。そして意見 1 にもある思いがけない発見の機会が高まる工夫が必要となる。

むすび

人文科学系の多様な情報資源を共有化することを目的として諸機関のデータベースを統合的に検索できるシステムに、国立歴史民俗博物館のデータベースの中から選んだ資料の目録を収録した二つのデータベースを適用する方法について検討し、実証システムにより評価を行った。

比較的汎用なメタデータであることから採用した Dublin Core に、資料の目録情報からなるデー

タ項目を一応の根拠をもって対応付けができることを示した。

しかし、資料の形状、状態、材質といった実際のものが持つ属性では、その意味の捉え方によって、対応付ける先にゆれが残る。また、ユーザインタフェースとしてエレメントの名称を直接的に検索語の入力欄に表記したのでは、実際に対応付けられているデータの内容との解離が起こる。エレメントの定義や名称を適正化する必要があることが実証実験を通じた利用者の観点からの評価により明らかとなった。検索結果の表示方法の評価より、表示するデータ項目や名称の表示方法の検討が必要であり、統合検索と個別検索の役割の整理が重要な課題であることが示された。また、多数のデータベースを対象とすることから、検索結果の表示と絞り込みに関するユーザインタフェースの面で新たな考慮が求められることが明らかとなった。

今回の実証実験で明確となった課題は、多様な研究資源を多様な視点から情報化したデータベースを統合化しようとするところから派生するものであり、本質的な解決が必要である。今回は対象としなかったデータベースも含めて総合的な検討を行い、情報資源共有化のあり方について研究を進める必要がある。

謝辞

本研究は、総合研究大学院大学共同研究「文化科学研究分野における情報資源共有化のためのコラボレーション研究」の一環として行った。この研究代表者である総合研究大学院大学の安永尚志教授（国文学研究資料館）をはじめ貴重なご意見と知見を与えて頂いた研究グループの各位に深謝する。また、システムの構築に当たり支援をいただいた共同研究グループの一員でもある国文学研究資料館の原正一郎助教授に厚く感謝する。

参考文献

- [1]——村田良二，“博物館情報の相互運用と CIDOC CRM の役割,” 情報処理学会シンポジウム論文集, vol.2002, No.13. pp.39-42 (2002).
- [2]——山本泰則, 中川 隆, “Z39.50 CIMI プロファイルをもちいた民族学標本資料の情報共有に向けて,” 情報処理学会シンポジウム論文集, vol.2003, No.21. pp.9-16 (2003).
- [3]——五島敏芳, “日本における記録史料記述 EAD/XML 化の実践—記録史料管理と EAD 利用をめぐって—,” 情報処理学会シンポジウム論文集, vol.2002, No.13. pp.217-224 (2002).
- [4]——牟田晶平: “インターネット上での国際的な歴史記録の共有を目指して—アジア歴史資料センターの事例,” 情報処理学会シンポジウム論文集, vol.2002, No.13. pp.201-208 (2002).
- [5]——原正一郎, 柴山 守, 安永尚志, “メタデータによるデータベースの機関間連携の実現—人文科学データ共有化のための標準化—,” 情報処理学会シンポジウム論文集, vol.2003, No.21. pp.17-22 (2003).
- [6]——安永尚志 編, 2003年総合研究大学院大学共同研究プロジェクト「文化科学研究分野における情報資源共有化のためのコラボレーション研究」第1回報告書, Jan. 2005.
- [7]——“DCMI Metadata Terms”, Dublin Core Metadata Initiative, Jun. 2004, <http://dublincore.org/documents/dcmi-terms/>
- [8]——山本泰則, 原正一郎, 柴山 守, 安達文夫, 合庭 惇, 安永尚志, “Dublin Core メタデータと Z39.50 にもとづく人文科学系データベースの統合検索に関する実証実験,” 情報処理学会シンポジウム論文集, vol. 2004, No.17. pp.199-205 (2004).
- [9]——“Guide to Best Practice: Dublin Core Version 1.1”, CIMI Consortium, Apl. 2000, http://www.cimi.org/public_docs/meta_bestprac_v1_1_

210400.pdf

[10]——“DCMI Type Vocabulary,” Dublin Core Metadata Initiative, Jun. 2004,
<http://dublincore.org/documents/dcmi-type-vocabulary/>

[11]——“Dublin Core Qualifiers”, Dublin Core Metadata Initiative, Jul. 2000,
<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

(国立歴史民俗博物館研究部)

(2005年3月31日受理, 2005年7月15日審査終了)

Museum Materials for the Sharing of Information Resources : Database Mapping and Its Evaluation

ADACHI Fumio, SUZUKI Takuji, KOJIMA Michihiro, TAKAHASHI Kazuki

In the field of the humanities, various databases have been compiled and made available to the public by means of numerous networks. Making it possible to search all of these databases simultaneously would make it possible for users to search without being aware of the location or operation procedures of individual databases.

Key institutions of the School of Cultural and Social Studies at the Graduate University for Advanced Studies and several universities are conducting joint research on systems that will enable global searching of all databases run by these institutions. In order to achieve this global search capability, the Dublin Core metadata standards have been chosen as one type of standard window onto these databases. The research group has also studied methods of mapping data items from Rekihaku's "General Catalog Database" and the "Medieval-Monjo Database," and performed evaluations using trial systems.

Data items formed from indexed information on materials can, with some tentative basis, be mapped to Dublin Core metadata, which was primarily established for the description of online resources. However, there remains some inconsistency in the mapping due to the interpretation of the meaning of the attributes of actual things, such as the shape, state and quality of materials. Also, transcription of element names directly into the search criteria entry fields of the user interface can result in cases where they are very different from the contents of data that was actually mapped. Verification tests confirmed that there is a need to moderate the definitions and names of elements. An evaluation of methods of displaying search results has shown that a study must be made of the data items to be displayed and the methods of displaying names, and that reorganization of the roles of global searching and individual database searching is an issue that needs to be addressed. Because the system deals with numerous databases, it also became evident that further consideration needs to be given to the user interface with regard to the display and refinement of search results.