



震災デジタルアーカイブにおけるメタデータの管理 と検索 : 資料中の構成要素の取扱いを中心に

渡邊, 隆弘

(Citation)

図書館界, 56(1):14-29

(Issue Date)

2004-05

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90000337>



震災デジタルアーカイブにおけるメタデータの管理と検索： 資料中の構成要素の取扱いを中心に

渡邊 隆弘

(わたなべ たかひろ 神戸大学附属図書館)

Metadata management and retrieval of Digital Archive for Hanshin-Awaji Earthquake Materials, by WATANABE Takahiro.

電子図書館におけるメタデータの有効性を高めるには、その意味的側面や構文的側面に加えて、データ作成単位や資料構造の表現についての検討が必要である。

神戸大学附属図書館の「震災デジタルアーカイブ」は、資料中の各構成要素をメタデータの作成単位とし、資料を構成要素群からなる階層構造ととらえる組織化手法に特色がある。本稿では、資料中の構成要素の取扱いに焦点を絞って、本アーカイブにおけるメタデータ作成・蓄積・検索に関わる諸問題を考察する。

0. はじめに

「電子図書館」活動の一つの方向性として、多くの機関で図書館所蔵資料のデジタル化・ネットワーク公開が実践されているが、その際に情報組織化や検索手法の問題が十分顧慮されているとはいえない。一次情報がデジタル化される電子図書館においても、全文検索の信頼性やマルチメディア情報の存在を考慮すれば、適切な「目録」、すなわちメタデータの管理が求められるということは概ねの共通認識であろうが、どのようなメタデータが求められるのか、これまでの図書館目録とどこが同じでどこが違うのか、といった具体的検討の蓄積は乏しいように思われる。多くの機関での出発点は、比較的少数の「お宝」資料（歴史的貴重資料）に対して原資料の再現性を重視した提供を行うことであったから、一定の「量」を前提とする組織化や検索の問題が後回しになってきたのもやむを得ない面がある。しかし、実験期から実用期へと発展し、国立国会図書館の「近代デジタルライブラリー」¹のような巨大なサービスも出現した現在、伝統的な情報（資料）組織化論の基礎にたって新たな可能性を探る作業が求められているものと考えられる。

筆者は既発表論文²で、電子図書館におけるメタ

データ記述規則を考える枠組みとして、「記述の単位と構造表現」「意味的側面」「構文的側面」の3側面を設定して論じた。このうち、メタデータにどのようなデータ項目を設定して何を記録するかという「意味的側面」と、どのような文法・構文でデータ表現を行うかという「構文的側面」に比べて、データ作成単位（目録規則の用語でいうところの書誌単位）や資料構造の表現についてはこれまで言及されることが少なかった。しかしながら、印刷資料のような確固とした物理的実体を持ちえず、またそれゆえに情報のより柔軟な提供・利用可という能性を持つ電子図書館資料においては、データ作成単位についてもこれまでの目録作成とは異なった取扱いが求められよう。様々な情報が組み合わさって一つの資料が構成されていることは非常に多いが、資料がデジタル化されていれば、個別の利用者要求に対してその都度必要な部分を検索・抽出するアクセスがシステム設計次第で比較的容易に提供できると考えられる。ただし、それを実現するためには、抽出すべき「部分」をどう認識するか、ひいては様々な情報からなる資料の「構造」をどう把握するか、が十分に検討されなければならないのである。

本稿では、神戸大学附属図書館の「震災デジタルアーカイブ」³におけるメタデータの作成・蓄積・

検索について、資料中の構成要素の取扱いに焦点を絞って考察する。阪神・淡路大震災関係資料を対象とする震災デジタルアーカイブは、大量の現代資料を対象とするという点で現時点ではまだ珍しい電子図書館事業であり、メタデータ構築に独自性の強い組織化手法を取っている。また、約5年の運用の中で、データ蓄積・検索方法の問題点を考察して一定の改善をはかってきたシステムでもある。以下、本稿ではまず先行例と問題の所在について述べ(1章)、震災デジタルアーカイブの対象資料・システム構成・検索機能を述べた(2章)うえで、資料を構成要素からなる階層構造ととらえるメタデータ管理方法(3章)と階層構造における検索の問題(4章)について詳述する。なお、筆者は本アーカイブの設計・運用に携わってきた経験を持っているが、本稿に示された整理・評価等は筆者個人の見解であることをおことわりしておく。

1. 資料中の構成要素の取扱い—問題の所在

資料の構成要素に踏み込んだ情報の記録は、これまでの図書館目録ではあまり積極的に運用されてこなかったが、その必要性は指摘されてきた。例えば古川は、複数の著作を含む資料が増加するなかで、「物理単位を対象とする記録」に加えて「著作を対象とする記録」が必要であり、構成部分に関する目録規則の整備を提唱している⁴。また主にOPACの主題検索改善という目的から、書誌情報に加えて目次情報を利用することでアクセスポイントの増強をはかるという視点もかなり以前から提言されている⁵。さらに、90年代に登場したネットワーク情報資源では様々な粒度の資料が混在するのが特徴の一つであるが、永田は早い時期に、電子的情報オブジェクトをも対象とする「次代のOPACにおける統合的な目録データベース」では図書を想定した「基礎書誌レベル」では不足であり、「下位書誌レベルとして扱われるもの」を含むさまざまなレベルのメタデータを統合する必要性を指摘している⁶。

また目録規則においても全く無視されてきたわけではない。『英米目録規則第2版』(AACR2)⁷では13章に短いながら「分出」(Analysis)の規定があり、『日本目録規則1987年版』(NCR87)⁸でも具体的規則には乏しいものの、基礎書誌単位の下位に構成書誌単位が設けられている。また、ISBD(国際標準書誌記述)においても(ガイドラインレベル

ではあるが)構成部分を扱うISBD(CP)が1988年に発表されている⁹。

一方実践面では、カード目録時代に分出は低調であったが、目録電算化の進展とともに、内容注記を充実させて1図書に収録された個々の著作からの検索を可能にすることが、かなりの程度行われるようになってきた。また、最近では目録規則に則った書誌情報とともに図書の目次情報を検索可能とする実用システムが登場してきている¹⁰。さらに電子図書館サービスにおいても、前述の「近代デジタルライブラリー」が目次入力を検索に役立てている。

これらの目録規則や実践例には、以下の2つの問題点があるように思われる。一つは、目録規則というより実践における問題点であるが、検索精度の問題への考慮がなされていないことである。内容注記や目次情報を書誌情報中に羅列的に記述するのが一般であり、再現率の向上という効果は大きい、一方でヒットした資料中のどこに求める情報があるのかを明確に指し示すことが難しい、また複数キーワードで論理演算を行った場合に資料中でかけはなれた位置にあるアクセスポイントがヒットして精度を下げる、といった問題が生じている。

もう一つは、資料構造のとらえ方が単純であり「資料全体—構成要素」という2層構造の表現にとどまっていることである¹¹。図書の目次を想起してみればわかるように、より複雑な構造を持つ資料もごく一般的に存在し、2層構造では資料を十分に表現したことにはならない。もっとも資料中の構成要素といっても、「著作」と呼べるような完結的な情報単位を対象とするなら、物理媒体「資料」と構成「著作」の関係がとらえられれば十分に実用であるともいえる。しかし一次情報がデジタル化された電子資料においては、「ファイル」などの媒体の単位はより便宜的・可変的な性格を持っており、資料の必要部分を検索・抽出するアクセスができるとともに、(一意に定められた線形的なアクセスだけでなく)その時々必要に応じた切り口で情報を提示できるところに、強みがある。この強みを十分に生かすには、著作・記事といった完結的な情報単位だけでなく、アクセス可能性が考えられればさらに細かな単位(記事中の章・節や図表など)を対象としうる枠組みが求められると考える¹²。

上記2点の問題について、資料中の構成要素がおりなす構造を用いた検索機能の注目すべきものとし

て、長尾らによる電子図書館実験システム”Ariadne”で行われた「階層構造検索」がある¹³。Ariadneでは図書の目次を利用し、目次に示された章・節・項のタイトルを階層構造としてとらえたうえで、「哲学＞人工知能」（哲学の観点から人工知能を論じたもの）というように上下の出現順序を指定して検索することにより精度の向上をはかっている。

2. 「震災文庫」と震災デジタルアーカイブ

本章では、アーカイブの母体となる「震災文庫」とその収集資料の特色、震災デジタルアーカイブの歩みとシステム構成、次いで検索機能の特色について述べる。

2. 1. 震災文庫とその収集資料¹⁴

震災デジタルアーカイブは、神戸大学附属図書館に開設されている「震災文庫」（阪神・淡路大震災関係資料文庫）を母体としている。震災文庫の資料収集は震災4ヶ月後の1995年5月に開始され、同年10月30日から同大学人文・社会科学系図書館の一角で一般公開を開始している。震災資料の収集・保存活動は被災地を中心とする諸機関で行われているが、公開されているコレクションとしては最大規模のものである。震災後9年を経た現在も収集活動を継続しており、2004年1月末現在で約37,000件（約20,000タイトル）の資料が所蔵されている。

震災文庫は開設当初から、媒体・形態を問わず、関係資料を網羅的に収集することを目指している。震災に関連があれば、公刊された図書・雑誌だけでなく非市販の報告書類や広報誌紙類も、パンフレット・レジュメ・チラシ・文書・ビラ・ポスターなどといった一時的な情報伝達を想定した形態のものも、また写真・ビデオ・録音など（市民による個人的な撮影を含む）様々なメディアを用いたものをも含む、あらゆる資料を収集している¹⁵（表1に媒体別の登録数を示す）。さらに、当然のことながら、雑誌論文なども含め資料の一部に震災への言及を含むものも収集対象である。阪神大震災に関する記録の大きな特徴は、行政機関によるもの以外に様々な組織や個人によるものが、震災直後から今日にいたるまで膨大に出されていることであり、活発な学術研究活動や各種機関における情報公開の趨勢、震災直後からの収集活動等もあいまって、大量の資料が収集されている。

通常の図書館資料と比べた場合の震災文庫資料の特色は、一言でいうと著しい「多様性」である。もう少し敷衍してみると、これには「資料媒体の多様性」「資料となる単位が多様性」「内容・形式の多様性」という諸側面がある。

表1. 資料区分別所蔵資料数（2004.2.8現在）

雑誌など複数の物理的単位からなる資料について、各号を1冊ずつ数えたのが「件」、同一タイトルは1つと数えたのが「タイトル」（ただし、データ管理上の制約から、「新聞・広報紙」の一部に、号ごとにカウントされていないものがある。）

	件数（タイトル数）
図書資料	5,298(4,790)
雑誌資料	10,802(2,817)
新聞・広報紙資料	11,132(1,894)
パンフレット資料	4,510(4,472)
一枚もの資料	5,209(5,209)
写真資料	82(81)
地図資料	132(126)
映像資料	189(162)
音声資料	67(61)
コンピュータ資料	66(66)
総合計	37,487(19,678)

資料媒体の多様性は表1でも明らかである。特に注目すべきことは、「新聞・広報紙資料」「パンフレット資料」「一枚もの資料」の多さである。図書館資料の多様化・マルチメディア化という点、映像・音声といった視聴覚資料や電子資料に目を向けられることが多いが、震災文庫では「紙媒体の非図書資料」と呼ぶべきものが多様性の一つの中心になっている。

次に、「資料となる単位が多様性」である。「震災」という切り口で資料収集を行なうと、一冊の図書・雑誌といった発行時の物理単位ではなく、その一部分だけが目的に合うことがしばしばあるため、抜刷・抜粋・切抜といった形態で保持されている資料が相当数に及んでいる。また、完全形の資料が所蔵されていても、実際には一部分だけが震災関連情報である場合も多い。図書館における通常の情報組織化作業では、出版された物理単位を固定的に扱うことを前提としており、こうした事態に対応する規則・ノウハウは蓄積されていない。

加えて「内容・形式の多様性」である。震災資料は記録を主目的としたものにとどまらず、救援・復

興の活動過程の中で生まれた様々な資料が大きな部分を占めているため、千差万別の内容を含んでいる。また、資料の作り手が多岐に渡り、いわゆる「編集」作業のレベルが一定しないため、章節構造や文章の書き方、図表の扱い方など構成・形式面でも差異が甚だしい。

このような多様性は、特定のテーマについて十全に資料収集しようとすれば多かれ少なかれ生じることであるが、震災資料においては、一般市民を含めて情報の発信主体が広いことから、特にその程度が甚だしいといえる。

2. 2. 震災デジタルアーカイブの歩みとシステム構成

震災文庫では当初からインターネットでの情報提供を方針の一つとしており、公開開始にあわせて目録検索等の機能からなる「震災文庫ホームページ」を開設した¹⁶。また 1998 年には、チラシ・ポスターなどの「一枚もの資料」について、写真撮影した資料画像のネットワーク公開を開始している¹⁷。

その後、神戸大学に「電子図書館化推進経費」が予算措置され、1999 年に「神戸大学電子図書館システム」が稼働をみた¹⁸。電子図書館事業には様々な方向性があるが、神戸大学電子図書館システムの最大の力点は所蔵資料をデジタル化して広く全国・世界に情報発信する「電子アーカイブ」事業に置かれ、「震災デジタルアーカイブ」はそのコンテンツの一つの柱に位置づけられた。現在までの約 5 年間、メタデータの整備と、一次情報デジタル化の推進の両側面からコンテンツ整備が行われている。メタデータ整備においては、新着資料の入力に加えて、詳細情報の遡及的入力を行っており、また一次情報デジタル化の有無にかかわらず統一基準でメタデータを作成している。一次情報のデジタル化は、著作権許諾と入力コストという両面の問題から進捗には限界があるが、2001 年ころから本格的な軌道に乗り出し、表 2 のような現状にある。

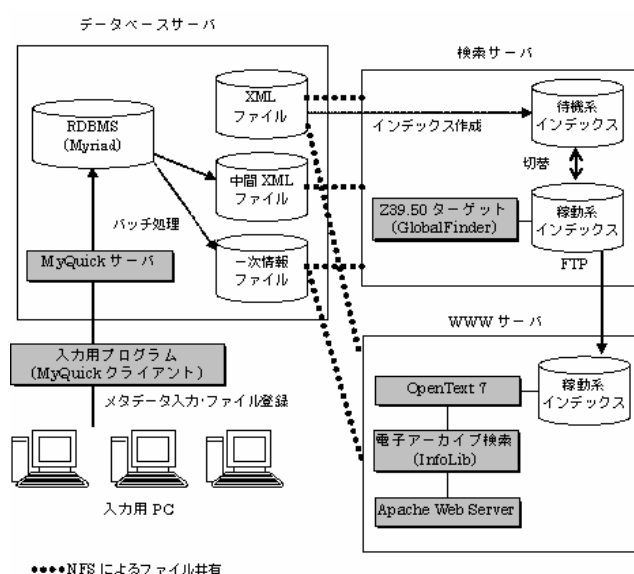
システム面では、1999 年 2 月よりメタデータ及び一次情報の登録・蓄積・検索を制御する統合システム（NTT 西日本株式会社の開発による）が導入された。これは震災資料のみを対象とするのではなく、その他のアーカイブをも統一的に扱う汎用システムである。その後、2003 年 2 月にシステム更新を行い、新システム（インフォコム株式会社の開発による）

に移行した。以下本稿では、前者（1999-2003. 1）を「前世代システム」、後者（2003. 2-）を「現行システム」と呼ぶ。図 1 に現行システムのシステム構成を示す。メタデータと一次情報は PC ベースで随時入力され、データベースサーバ（サーバの OS はすべて UNIX）に蓄積される。この時点のデータ管理は RDBMS（リレーショナルデータベース）を用いている。検索系では、国内外で広く使われている全文検索エンジン OpenText7 を使用している。OpenText7 は検索対象データを単語や分節に分割する形態素解析は行わず、単純に文字列の出現を検知する検索エンジンである。OpenText7 は XML 形式データを標準入力と想定しているため、メタデータは RDBMS から日次に XML 形式への変換を施して格納している。

表 2. 一次情報デジタル化資料（2004 年末現在）

	タイトル数	総データ量
図書	173	約 28,000 ページ
雑誌・広報紙・新聞	56	190 号分
パンフ・文書集	397	
一枚もの	2,855	
写真	12	約 24,000 枚
音声	2	約 110 分
動画	4	約 300 分
統計データファイル	1	
合計	3,500	

図 1. システム構成図（2003. 2～）



2. 3. 震災デジタルアーカイブの検索機能

機能の具体的説明に入る前に、情報の検索という観点から震災デジタルアーカイブに求められる要件について述べておく。前述のように、震災文庫資料の特色は「資料媒体」「資料となる単位」「内容・形式」といった側面における「多様性」にある。この多様性をどのようにコントロールすれば利用者に適切に提供できるのか、が震災デジタルアーカイブに課せられた目標であるといえる。

さらに、資料自体の特性とは次元が違うが、利用者の情報要求の多様性という問題がある。もちろんどのような分野においても多様な情報要求があるが、従来図書館が作ってきた目録は、各主題の様々な二次資料（書誌・索引・事典など）との連携で最終的に利用者を資料に導いているものであり、全情報要求を一手に引き受けるものではない。ところが、震災関係情報には網羅的な記事索引や二次情報データベースが存在していないため、震災デジタルアーカイブには震災に関する多様な情報要求のすべてを解決するレファレンスデータベースとしての役割が期待されることとなる。

震災デジタルアーカイブの検索機能は、「デジタルアーカイブ検索」の名称で提供されている。検索画面はシンプルな作りの「簡易検索」と複雑な代わりにきめ細かな設定のできる「詳細検索」の2種類が用意されている（このこと自体は目新しくないので詳しくは述べない）。様々な媒体からなる震災資料に対して、その全体を対象とするほかに、特定の形態の資料のみ（たとえば「写真」のみ）に限った検索も可能である。

図2は検索結果の一覧表示画面であり、タイトルなど資料の書誌情報が表示されている。ここで、本システムの特徴として、次の2点があげられる。

- (1) メタデータを介した検索であること
内容全文がテキストデータ化されている資料もあるが、検索の基本は資料に関して作成されたメタデータに対して行われる。
- (2) 資料中の一部分を独立して検索できること
資料1点の全体ではなく、その中の記事・章節、あるいは1枚の写真といった一部分が検索され、独立して結果表示される。これにより、入力したキーワードに合致する情報を含んだ箇所をピンポイント検索することが可能である。結果として、一覧表示されるデータは、図書1冊であったり、そのごく一部であったり、表示される情報単位(粒度)が一定しなくなるので、各項にデータの種別を（媒体の種類とともに）示している。

図3は一覧表示から1件を選んだ詳細表示画面である。ウィンドウが左右のフレームに分割されているが、右フレームが選ばれたデータの詳細情報である。この場合は写真集の中の1枚の写真のメタデータである。それに対して左フレームは、この写真集から作られたメタデータを一括して一覧する表示である。各写真に対するデータの他に、資料全体のデータや、その中間的なレベルのデータなどもある。それらを一括して表示すると一個のツリー構造になるので（表示上はインデクションでレベルを示している）「階層型表示」と命名しているが、別の言い方をすれば資料構造の全体を示す「目次」にあたる。また、右フレームの上部には一次情報へのリンクが示され、クリックにより一次情報の閲覧ができる。

以上のように、「電子アーカイブ検索」ではキーワードによる資料検索から一次情報の閲覧までが、一つの流れで行えるようになっている。

図2. 検索結果一覧表示画面

神戸大学 電子図書館システム -- 検索結果一覧 -- Microsoft Internet Explorer			
アドレス http://www.lib.kobe-u.ac.jp/das/SearchServlet			
92 / 2	復興再開発事業における合意形成の過程：六甲道駅南 / 有光友興 仲川邦俊 三浦哲之 震災/図書/記事/論文 .jpn 学芸出版社, 1998.2, p.59-65 (震災復興が教えるまちづくりの将来：阪神・淡路大震災/「復興まちづくり報告'97」を全国に伝えたい：はじめに/復興の現場から/復興再開発・共同化等) 請求記号：震災-4-528		00008313
93 / 1	復興市街地再開発事業：JR六甲道駅南 JR新長田駅南 進ちょく状況“東西格差”：地権者との交渉に時間 震災/雑誌/記事/論文 .jpn 兵庫県震災復興研究センター, 2002.3, p.16 (震災研究センター = Hyogo Research Center for Quake Restoration, no.47-no.99 / no.79) 請求記号：震災-4-z24		00317578
94 / 2	北東側から見た復旧当日の六甲道駅：上りコンテナ列車が通過中 [PRO Photo CD] / 進藤 裕之 震災/写真/写真 .und 神戸大学附属図書館(制作), 1995.12, 1枚(フレーム4-73) (阪神大震災記録写真集：ある大学院生の目を通して/1995年4月1日(土)：JR完全復旧：瀬～住吉間開通) 請求記号：震災-1-95		00225189
95 / 2	まちづくり素案の考え方：JR六甲道駅前・永手町5丁目まちづくり協議会意見交換会第2部資料 震災/パンフ/タイトル .jpn JR六甲道駅前・永手町5丁目まちづくり協議会, 1996.2, 9p 請求記号：震災-4-v183/2,v184/2,v254/2		00004925
96 / 1	まちづくり提案(街区別施設配置計画)を終えて：六甲道駅南地区震災復興第一種市街地再開発事業の誕生。第2回 / 若井吉昭		00332829

図3. メタデータ詳細表示画面

神戸大学 電子図書館システム -- メタデータ詳細表示 -- Microsoft Internet Explorer

アドレス http://www.lib.kobe-u.ac.jp/das/jsp/DetailView.jsp?LANG=JA&METAID=00225189

階層型表示

阪神大震災記録写真集：ある大学院生の目を通して

「阪神大震災記録写真集」緒言

それまでの神戸

JR三宮駅付近からフラワーロードを望む

神戸市風景

震災前のJR六甲道駅

1995年1月17日(火)：地震当日

炎上中の神戸市中心部

炎上中のJR六甲道駅周辺

全壊した六甲道駅の北側出入口付近

JR六甲道駅の西側の高架下：大学院法学研究

六甲道駅の南側出入口付近

明石：自宅の被害を中心に

傾いた顔：明石市の神戸市に隣接する地域

LPプレーヤーは転落したが、ステレオは落下

東向きに倒れた書架：その奥に、冷蔵庫は北

●一次情報

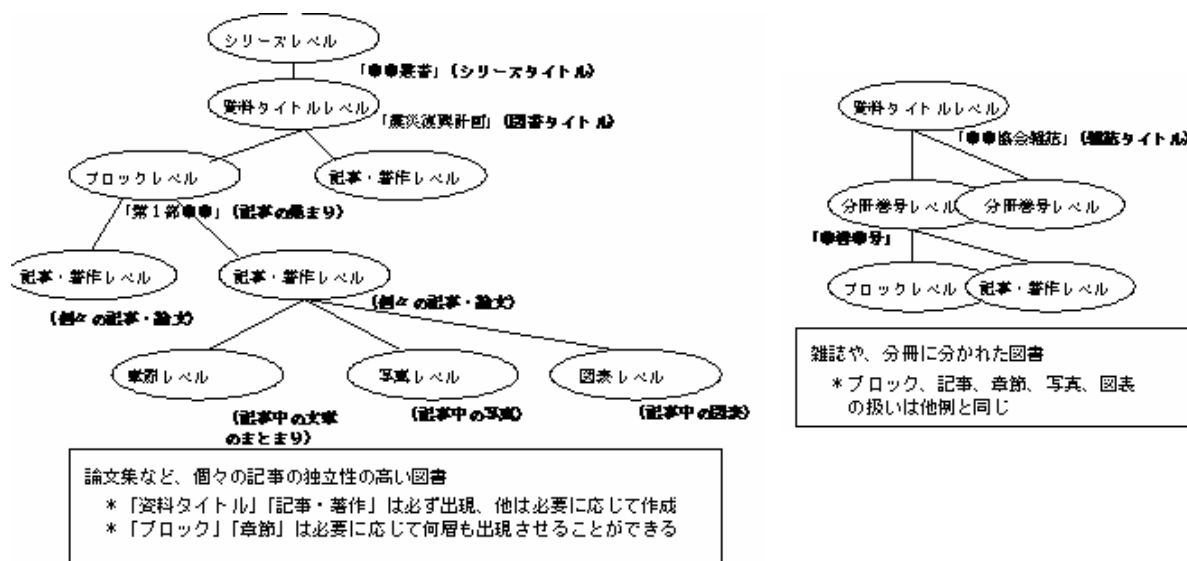
※HTMLには新聞記事などの本文データが含まれています。
サムネイル [1]

表示用画像 [1]

●メタデータ

タイトル	北東側から見た復旧当日の六甲道駅：上りコンテナ列車が通過中
資料種別	PRO Photo CD
著者など	進藤 裕之 撮影 神戸大学大学院法学研究科
上位タイトル	阪神大震災記録写真集：ある大学院生の目を通して/1995年4月1日(土)：JR完全復旧：瀬～住吉間開通
ページ数など	1枚(フレーム4-73)
出版地	神戸
出版者	神戸大学附属図書館(制作)
出版年月	1995.12
メタデータID	00225189
旧メタデータID	00093922
リソースID	震災/写真/写真
本文の言語	und
データ登録日	20000216
所在	人文・社会科学系図書館震災文庫
請求記号	震災-1-95

図4. 重層的なメタデータ作成の模式例



3. 震災デジタルアーカイブにおけるメタデータ¹⁹の作成単位

本章では、2.3 で述べた検索機能を実現するためにとられているメタデータ管理手法について述べる。

3. 1. 構成要素ごとのメタデータ作成

入力キーワードに合致した箇所をピンポイント的に検索・表示するため、資料中の構成要素を独立して操作対象とする。すなわち、個々の構成要素ごとにそれぞれ独立したメタデータセットを作成し、それぞれに対応した単位の一次情報とリンクさせている。ここで「独立」とは他の要素と明確に識別可能なデータ構造を意味し、必ずしも独立した「レコード」を作成するということではない（構文的側面は3.3で後述する）。

図4にメタデータ作成の模式例を示した。「構成要素ごと」とは、例えばある雑誌中の論文・記事ごとに独立してメタデータセットを作成するということであるが、同時に誌名や巻号から収録論文を通覧するアクセスも十分に考えられるため、そうしたレベルでも別途レコードを作成する。つまり、雑誌資料であれば一つの「資料タイトル（誌名）レベル」、冊数分の「分冊巻号（各号）レベル」、さらに個々の「記事・著作レベル」がそれぞれ作成される。さらに検索上の必要に応じて、「ブロックレベル（“特集”のように複数記事をまとめたレベル）」や「章・節レベル（1記事中の部分レベル）」が作成され、時には掲載され

た写真・図版等を「写真レベル」「図表レベル」などの形で扱うことも行われる。すなわち、一つの資料から、必要と思われる限りにおいて、多数のメタデータセットが重層的に作成されうることとなっている。もちろんこうしたデータ作成は多大な労力を要するもので、漸及的な入力計画が進行中である。

作成された個々のレベルのメタデータセットは、検索システム上、等価に扱われ、検索結果一覧やメタデータ詳細表示において独立して表示される。そのため個々のメタデータは、表示を自己完結させるだけのデータ項目を持っている。例えば、出版年月等の出版・発行に関わるデータは、物理的に刊行された単位（上述の雑誌の例では「分冊巻号レベル」）に帰属するものであるが、本アーカイブでは、収録された個々の記事レベルのメタデータにおいても、同じ値を保持している。

3. 2. 「リソース種別」による管理

表3. リソース種別

アーカイブ種別	資料種別	エレメント種別
01 震災文庫	01 図書資料	01 シリーズレベル
.....	02 雑誌資料	02 資料タイトルレベル
その他アーカイブ名称	03 新聞・広報紙資料	03 分冊巻号レベル
	04 バンフレット資料	04 ブロックレベル
	05 一枚ものの資料	05 記事・著作レベル
	06 地図資料	06 章・節レベル
	07 写真資料	07 写真レベル
	08 映像資料	08 地図レベル
	09 音声資料	09 図表レベル
	10 コンピュータ資料	

レベルごとにメタデータを作成すると、データベース中には様々な情報単位のメタデータセットが混在することになる。これらの統合検索には大きな利点があるが、一方で検索結果の単位がばらばらでわかりにくいなどの問題が発生する。

このため本アーカイブでは、検索結果上で「震災／図書／記事・論文」のように、それが「どういう資料のどういうレベルのものか」を明示しているが（図2を参照）、この情報を各メタデータに付与されたコード情報「リソース種別」によって管理している。リソース種別は表3に示したように3種類のコードよりなる。ただし、資料群の種別である「アーカイブ種別」は震災資料ではすべて同一であり（本システムは他の資料群をも扱う汎用システムであるために設けられている）、「資料種別」と「エレメント種別」がメタデータの種類を表している。

「資料種別」は、図書資料・一枚ものの資料・映像資料といった、対象資料の媒体・形態を表すコードである。当該資料全体をとらえての区分であり、同一資料から作られた各レベルのメタデータの資料種別は同じになる。なお、資料種別には、「図書」や「コンピュータ（ファイル）」のように表現された情報をどう格納するかという外形式的なものと、「写真」や「地図」のように情報の表現方式に踏み込んだ内形式的なものが両方含まれている。図書として発行された地図帳は「図書」か「地図」かというように両者の衝突が時に生じるが、この場合は内形式を優先させており、シートの地図・地図帳・電子化された地図はいずれも「地図資料」と扱われる。

「エレメント種別」は、当該メタデータの構造上のレベルを表すコードであり、図4の「〇〇レベル」に対応する。これによって、1資料から重層的に作成されたメタデータ群の中で、各メタデータがどのような位置付けにあるのかの情報が得られる。なお、厳密にいうと、「エレメント種別」には、「シリーズ」「分冊巻号」「章節」のような構造上のレベルのみを表すものと、「写真」「図表」のようにレベルというより「当該構成部分データの資料種別」ともいえるべきものが混在しており、理論上は必ずしも明快でない。両者とも重要な情報であり、分割してリソース種別を4コード構成にすることも考えられるが、構造が複雑になりすぎるという判断から現行の実装となっている。

各メタデータセットをリソース種別（資料種別と

エレメント種別）で管理することによって、検索システム上、次の諸機能が実現されている。

- ① 検索結果として提示されたメタデータの資料媒体や情報単位を明確に提示する。
- ② 各リソース種別に最適なデータ項目設定や画面表示を行う²⁰
- ③ あらかじめ検索対象を特定種類のメタデータに絞り込むことができる。

このうち、③の検索対象指定については、資料種別とエレメント種別の二段構えの構成をとることで、必要に応じて柔軟な対応ができる。「図書に限って検索（資料種別「図書」に限定）」「図書のタイトルに限って検索（資料種別「図書」、エレメント種別「資料タイトルレベル」に限定）」等のほか、「写真に限って検索（資料種別「写真」またはエレメント種別「写真レベル」に限定）」といった条件で、写真集などの資料と他の資料（図書など）中に掲載された写真とを横断検索させることも考えられる。

3. 3. 階層構造情報の管理（構文的側面）

各レベルのメタデータセットがばらばらに存在しているだけでは、構成部分への検索は可能であっても、資料の全体像がとらえられない。ピンポイントアクセスの一方で、図3の左フレームにあるような、資料の文脈・構造を再現する表示も必要である。資料の全体像を示すツリー構造情報をどのように表現するかは、メタデータの構文的側面の重要な要素である。

震災デジタルアーカイブの前世代システム（1999～2003.1）では、構成要素ごとに全く独立したメタデータレコードを作成しており、特にツリー構造を考慮したデータベース設計ではなかった。そのうえで、階層型表示を実現する仕掛けとして各メタデータレコードに、ツリー構造上での当該メタデータの位置を指し示すデータ項目「ツリー内階層パス情報」を持たせていた（図5）。資料構造の表現には、メタデータ間の上下関係を示す情報と、同位メタデータ内での順序関係を示す情報が必要であるが、「ツリー内階層パス情報」はその両情報を保持している。あるメタデータレコードと同じツリーに属する²¹メタデータ群を検索して、「ツリー内階層パス情報」でソート（並び替え）し、「階層パス情報」の長さを調べて適切なインデクションを施しながら順次表示していけば、当該資料の「階層型表示」が作成される

こととなる。理論上は無限・自由なツリー構造を表現可能である。

図5. 階層パス情報による構造表現

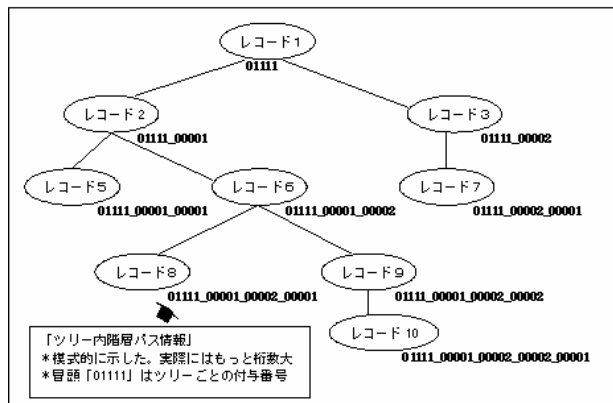


図6. XML 記述による階層構造表現

1<MATERIAL>	
2<S>	
3<NODE>メタデータセット1<NODE>	階層深度 1
4<S>	
5<NODE>メタデータセット2<NODE>	深度 2
6<S>	
7<NODE>メタデータセット5<NODE>	深度 3
8<S>	6行目に対応
9<S>	
10<NODE>メタデータセット6<NODE>	深度 3
11<S>	
12<NODE>メタデータセット8<NODE>	深度 4
13<S>	11行目に対応
14<S>	
15<NODE>メタデータセット9<NODE>	深度 4
16<S>	
17<NODE>メタデータセット10<NODE>	深度 5
18<S>	16行目に対応
19<S>	14行目に対応
20<S>	9行目に対応
21<S>	4行目に対応
22<S>	
23<NODE>メタデータセット3<NODE>	深度 2
24<S>	
25<NODE>メタデータセット7<NODE>	深度 3
26<S>	24行目に対応
27<S>	22行目に対応
28<S>	
29<NODE>メタデータセット4<NODE>	深度 2
30<S>	28行目に対応
31<S>	2行目に対応
32<MATERIAL>	

現行システム（2003.2～）では、4章で述べる検索上の問題を解決する必要等から、異なった手法をとっている。すなわち、1資料1レコードとし、図6のようにXML記述による構造表現を行う（図6は、図5と同じツリー構造を表現している）²²。図中の各「メタデータセット」には各構成要素の書誌情報（タイトル・著者など）がデータ要素ごとにタグを付してXML記述されており、資料の構造にもよるが、時には相当複雑・巨大なメタデータレコードになる。

しかし、本稿で検討しないメタデータセット内部の情報表現を捨象すると、資料構造の表現は極めて単純なタグで行われている。図6では<MATERIAL>、<S>、<NODE>という3種類のタグが現れているが、<MATERIAL>～</MATERIAL>は資料全体を、<NODE>～</NODE>は各メタデータセットを表現しており、ツリー構造は<S>～</S>のタグを組み合わせることのみで行われている。例えば9行目の<S>は図5の「レコード6」に対応しているが、対応する終了タグは20行目の</S>であり、その間には図5中の「レコード8,9,10」にあたる記述が含まれている²³。このように各メタデータセットに対応する<S>タグを階層関係に応じてネスト（入れ子）記述することで、無限・自由なツリー構造を表現することができる。

なお新旧どちらの方式をとるにせよ、このような階層構造情報を手作業による入力で管理するのは不可能である。本アーカイブの入力システムでは、Windows系OSのファイル管理に使われる「エクスプローラ」を模した特別なインターフェースで、入力者が視覚的に管理できるようになっている。

3. 4. 構成要素の扱いに関する独自性

本アーカイブでとっている手法の特色としては次の2点があげられる。

一つは、レベルごとに独立したメタデータセットを作成していることである。基礎書誌レベルの書誌レコード中に内容注記等の構成要素情報を記述してアクセスポイントを増加させる方法はよく用いられているが、多くは羅列的な記述で、構造化に乏しい。その結果、アクセスポイントの増加で構成要素情報からの検索は可能であっても、ヒットした資料中のどこに求める情報があるのかを明確に指し示すことが難しく、また複数キーワードで論理演算を行った場合に資料中でかけはなれた位置にあるアクセスポイントがヒットして精度を下げるといった問題が生じる。本アーカイブでは、各構成要素を明確に区別したデータ記述を行うことでピンポイントアクセスを可能にすると同時に、資料構造を示す情報（すなわち各構成要素間の関係情報）を明確に保持することで資料の構成・文脈に沿ったアクセスをも実現している。

もう一つの特色は、独立して扱われる構成要素が「著作」と呼ぶような完結的な情報単位（もしくははその集合）にとどまらず、著作中の章・節や文

章中の図表などの細かな単位についても、アクセス可能性に応じてメタデータセットが作成されうることである。前章で述べたように、震災情報を対象とする二次資料の乏しさから、震災デジタルアーカイブは多様な情報要求を直接受けとめなくてはならないが、その際、本来それ自体では完結していなかったような小さな情報が、問題解決の重要な手がかりになることがしばしばある。1. で述べた電子資料の強みを十分に生かすため、このような方針をとっている。

さらに、文書構造のモデルを特に作成せず、抽象的なツリー構造として扱っていることも特色の一つにあげられる。例えば XML による構造表現の際に〈章〉〈節〉のようなタグを設定し、ありうべき構造を DTD (文書型定義) であらかじめ定めておくことも考えられるが、本アーカイブでは各構成要素を等価に扱いタグ〈S〉のみで表現している。この方式により、資料の多様性に対応した柔軟な構造表現が可能である。

ところで XML 等のマークアップ言語はもともとテキスト文書に文書構造情報を付与するために作られており、文書構造情報を適切に付与することで単純な全文検索を超えたより精度の高いテキスト検索の実現をはかる方向性も考えられる。しかし、震災デジタルアーカイブでは一次情報をテキストデータとして保持する場合も、そうした格別の構造化は行っていない (一次情報の全文検索は可能であるが、文書構造を反映しない単純なものであり、メタデータ検索の補完的な位置付けでとらえている)。内容・形式面で多様性のはなはだしい震災資料においては、テキストに対して検索に有効なレベルの分節化をはかる DTD の作成は非現実的と考えられるからである。一方、資料の構成要素ごとに重層的にメタデータセットを作成して階層構造管理を行うことは、資料内の文書構造を保持して検索・表示に役立てることでもある。メタデータ作成は利用者のアクセス可能性を考慮した単位で行われるもので、テキスト中の重要な部分を明確化するような詳細な構造化とは異なるが、多様な資料を対象としながら、一次情報組織化について一定の代替機能を提供しているといえる。

4. 階層構造に関する検索の問題—論理積の扱い方を中心に

各構成要素を独立したメタデータセットと扱い、

資料構造をメタデータ群からなるツリー構造と捉える震災デジタルアーカイブのデータ管理手法は、検索面でいくつかの問題をもたらす。本章では、その問題点と改善方策について述べる。

4. 1. 前世代システムにおける検索方式

前述のように、前世代システムでは資料内の各構成要素を独立したメタデータレコードとしてデータベース構築がなされていた。特段の配慮をしなければ検索上も各構成要素は独立して扱われるが、それでは階層レベルをまたがった検索に問題が生じることが設計当初から意識された。

例) 神戸大学の被災状況 (資料タイトルレベル)

第1章 概説 (章・節レベル)

第2章 附属病院の活動 (章・節レベル)

このようなツリー構造の場合に、各レベルのメタデータが独立して取り扱われているとすると、「神戸大学 and 附属病院」といった論理積検索では、資料タイトルレベルのレコードも第2章のレコードも単独では条件を満たさず検索されない可能性がある。

「附属病院の活動」は孤立して存在しているわけではなく、あくまで「神戸大学の被災状況」の一つの章であるという資料構造の中に位置したものであるから、当然ながらそのような検索では使いものにならない。すなわち、1 資料から各階層レベルのメタデータセットが重層的に作成されているとき、上位のメタデータは下位の内容を包含している関係にあり、上位レベルの情報は下位にも継承されたものとして捉える必要がある²⁴。

一方で、階層をまたがった検索では、戻るメタデータセットに対する考慮も必要である。前記の考え方では資料タイトルレベル「神戸大学の被災状況」の情報を下位に継承して、章・節レベルをいわば「神戸大学の被災状況 — 附属病院の活動」と捉えようということになるが、ここで「神戸大学」とのみ検索した際には資料タイトルレベルのレコードがヒットすればよい。上位情報を継承したからといって各章・節レベルまですべて別個にヒットしてしまつては、一覧表示が混乱して正確なヒット件数もつかめなくなる。

以上の点を考慮して、前世代システムではメタデータ作成時に次の処理を施していた。

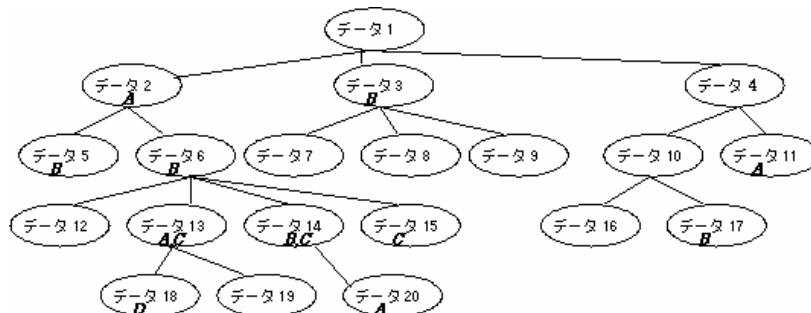
- ・あらかじめ、ツリーの上位階層にあたる部分のタイトルを自動的に下位レコードに埋め込んでおく。目録における「シリーズ」情報にあたるものである。
- ・特定の階層レベルに帰属するデータ項目を、下位レベルにも埋め込んでおく。例えば、出版者・出版年月等の出版事項は物理的に独立したレベルに本来帰属するものであり、「章・節レベル」

のような下位の構成要素には不要であるが、あえて同一項目を持たせる（当然同じ値が入る）。検索処理そのものは個々のメタデータレコードに対して独立的に行われるが、この際、

- ・ある検索条件に対して階層関係をなす複数のメタデータがヒットした場合は、最上位階層のものに絞り込んで表示する。

の処理を行っていた。

図 7. ツリー構造模式図



以上を、もう少し模式的に整理してみよう。図 7 のツリー構造（斜体で示された *A, B, C, D* がキーワード A, B, C, D の出現を表すとする）に対して「A and B」という論理積検索を行う場合、キーワード A がデータ 2 のタイトルに現れ、キーワード B がデータ 6 のタイトルに現われるなら、データ 6 は上位であるデータ 2 の情報をも継承しているので、この論理積検索にヒットすることとなる（前述の「神戸大学 and 附属病院」がこれにあたる）。このとき、検索結果として返戻されるのは、下位であるデータ 6 である。A, B の両方を満たす内容はデータ 6 まで限定できるので、ピンポイント的に目指す情報を示すには下位のメタデータを返戻するべきである。一方、データ 14 にもキーワード B が現れるがこの場合はどうか。この場合は、上位であるデータ 6 の内容がすでに「A and B」の論理積検索を満たしているから、データ 14 まで限定することはせず、データ 6 が返戻される。言い方を変えれば、データ 6 も 14 もともに「A and B」を満たすが、このように直列的な階層関係をなす複数データがともにヒットした場合は、上位のものを優先する（この仕組みで、上述の「最上位のものだけに絞込み」処理もできる）。なお、継承は下位へ下位へと順次行われるので、「データ 1-2-6-13-18」のように直列的に階層パスをたどれるメタデータ群のどこかに各キーワードが出現するな

ら、論理積検索にヒットしたものとみなされる（「B and D」でデータ 18 がヒットする）。キーワードが 3 個以上に増えた場合も同じ論理である。要するに、直列的な階層関係にあるメタデータ群を一体のものとみなして論理積検索を行うことを基本としているのである²⁵。

4. 2. 問題点の指摘

その後、震災デジタルアーカイブを題材にして情報工学研究者の立場から「デジタルアーカイブ」における効率的な検索手法を論じた研究が、依田・大月から神戸大学総合人間科学研究科のグループによって発表された²⁶。この研究では、資料構造を構成要素（依田らは「部分資料」と呼んでいる）からなるツリー構造として表現する手法を「種別やメディアを問わずに網羅的に資料が収容」されているようなアーカイブに対するアプローチとして有効であると評価したうえで、震災デジタルアーカイブの問題点を指摘し、新たな検索手法の提案を行っている。指摘された問題点は、次の 3 つに整理できる。

- (1) 階層レベルをまたがった検索に漏れが生じる場合があること
- (2) 直列的な階層関係だけの考慮では検索漏れが生じる場合があること
- (3) 返戻する構成要素の選定に不適切な場合があること

いずれも重要な指摘であり、以下、順に論じる。

階層レベルをまたがった検索について、依田らは震災デジタルアーカイブの検索方式を各メタデータセットへの「単純 AND 検索」であるとして、直列的關係にあるメタデータ群を一体に扱う「直列 AND 検索」が必要と主張した。筆者は前節に述べたように、前世代システムの検索処理自体は「単純 AND 検索」であるが、上位情報の下位継承を行うことによって依田らのいう「直列 AND 検索」を志向したものと位置づけている。ただし、上位からの情報継承には多分に不十分さがあった。システムの、上位タイトル（列）のみを下位メタデータに保持し、他の項目は入力時の運用（コピー入力）によってカバーする脆弱なものであり²⁷、完全な「直列 AND 検索」にはなっていなかったのも事実である。

さらに依田らは、「直列 AND 検索」だけでは検索漏れとなる場合があり、同一資料中での共出現をとらえる「並列 AND 検索」が有効な場合もあると主張した。再び図 7 に戻ると、直列 AND 検索で「A and B」の論理積検索を扱えば、返戻されるのはデータ 5,6 であり、「データ 2,3」「データ 11,17」のように直列的でない共出現は無視される。これは前世代システムの設計にあたって、意識的に行ったことである。例えば、「西宮市 and 避難所」といった検索を考えたときに、「西宮市」を含むメタデータの下位に「避難所」を含むメタデータがあれば（もしくはその逆）西宮市内の避難所について扱われている確率が高いが、そうでない場合は 1 資料中にたまたま共出現しているだけで、利用者の望む情報である確率は相当低いと推定される。1 資料中での共出現をすべて検出することは、検索精度の低下を招く危険性のほうが多いとの判断であった。これに対して依田らは、論理積検索に投入される複数キーワードには 2 つの種類があることを指摘した。一つは複数キーワードが修飾関係にある場合であり、もう一つは並立している場合である。前者は「西宮市 and 避難所」のように「A の B」（西宮市の避難所）と表現できる場合で、複合主題として構文表現できる関係とも言い換えられる。この場合は上述のように「直列 AND 検索」が有効である。一方後者は「西宮市 and 芦屋市」「図書館 and 博物館」のように「の」ではなく「A と B」と表現される場合であり、「並列 AND 検索」によって資料中の共出現を検出することが必

要である。この問題は前世代システム設計時には考慮の外にあった盲点であった。改めて検索語のログを採取すると並列と思われる入力は見られ（主題検索であるのか既知検索であるのかは判然としないが）、「並列 AND 検索」も提供の必要があることが認められた。

もう一つの問題点である返戻方式について、依田らは前世代システムの方式を「トップダウン方式」と呼び、それだけでは適切とはいえない返戻が起こる場合があるとして「ボトムアップ方式」の併用を提言した。前世代システムの方式は「上から見ていて、全キーワードが満たされる適切な位置を返す」という結果を生むものであり、「トップダウン」といえる。一方、例えば図 7 においてキーワード C で検索した場合、トップダウン方式ではデータ 13,14,15 の 3 つが返戻されるが、この場合データ 6 の下位メタデータの大部分がヒットしているのだから、データ 6 に集約して返戻するというのが「ボトムアップ方式」である。確かに、ピンポイント検索を目指すとはいえ、必要以上に下位データが多くヒットするのは問題であり、このような場合は集約が望ましい。依田らは集約判断の基準として、ヒットした下位メタデータの「数（いくつヒットしたか）」と「比率（全体の何割ヒットしたか）」をあげているが、階層構造の様相が資料によって多様であることを考えると比率による判断が適当である。

他にも階層構造に関わる検索の問題はいくつかある。一つは、各構成要素を全て等価に扱うことが適当かという問題である。実際には、資料タイトルや著作・記事といった完結的な情報単位とそれ以外など、レベルによって一般的な重みの差がある。すべてを等価に扱うと、階層型表示などでやや焦点がぼやけた表示となる場合がある。また、資料によって各構成要素間の関連性の強さには差がある。例えば雑誌では、他の資料ほど資料タイトル（誌名）と各記事との間に強い結びつきはなく、直列関係によるヒットでもややノイズが目立つことがある。並列 AND 検索（すなわち各記事間の関係）を導入するとなればさらに懸念される。

また、下位継承は直列 AND 検索に不可欠のものではあるが、時に適合すべきでないレコードを探してしまう現象が起こることがある。これは、「灘区・東灘区の復興状況」の下位に「灘区の復興」「東灘区の復興」の章があるといった場合に起こる。下位の

メタデータでは一つの区のみを扱っているのに、上位からの継承で「灘区の復興」にも「東灘区」というキーワードが合致することになってしまい、思わぬ検索結果が出る場合がある。「神戸市東部」－「灘区」「東灘区」のように下位を意味的に包含した用語が上位メタデータで使われている場合は問題がないが、下位項目を列挙している場合には誤ったキーワードを継承してしまうのである。これは語の意味論に関わる問題であって、今のところ有効な対処を考え難い。

4. 3. 現行システムにおける検索の改善²⁸

現行システム（2003.2～）の導入にあたっては、先に問題点を指摘した大月・依田らの研究グループの支援も仰いで開発を行った。3.3 で述べたデータ蓄積方法も大きな変更点であるが、検索面では4種類の論理積演算を実装することとなった。

①単純AND 演算子「AND」

同一メタデータセットに共出現

②直列AND 演算子「*」

直列関係をなすメタデータセット群に共出現

③親戚AND 演算子「&」

同一資料内の共出現（並列AND）

④兄弟AND 演算子「/」

同一の親をもつメタデータセット群に共出現

これに論理和（OR）、論理差（NOT）を加えた6種の論理演算子による検索式記述が可能である²⁹。論理積演算のデフォルト設定は「直列AND」としており、特に演算子をつけずに空白で区切って検索語を羅列した場合は、前世代システムと同様に直列的關係にあるメタデータ群を一体に扱った検索となる。ただし、上位情報の下位継承に不十分さのあった前世代システムとは異なり、確実な直列AND検索が可能な点に大きな改善がある。資料構造をXMLデータとして保持していても、検索エンジン（2.2で述べたように全文検索エンジンOpenText7を使用）の基本機能だけではこれらの論理積演算は行えず、個々のキーワードに対する検索結果リストを比較して適切にマージするアプリケーションを開発・実装している³⁰。

4つの論理積演算の中心は直列ANDと親戚AND（依田らの指摘した「並列AND」）であり、前節で述べたキーワード間の関係性の種類を考慮して、「（西宮市 & 芦屋市）*避難所」（西宮市「と」芦

屋市「の」避難所）といった組合せにより適切な検索が可能である。ところで、直列AND検索においては4.1.に述べたように入力キーワードの全てを満たすレベルのデータを返戻するが、親戚（並列）AND検索においては、各キーワードでヒットしたデータから上位に登って交わる位置のデータを返戻することとしている。例えば図7の構造をキーワードA,Bで親戚AND検索した場合、データ2,3をヒットと見なせばデータ1を返戻、データ11,17をヒットと見なせばデータ4を返戻、となる。ただ、そもそものA,Bの組合せを考えるか（例えばデータ3,11もヒットと見なせる）、上記操作後のデータ1,4の関係を再度考慮して集約するか否か、といった問題も生じ、直列ANDに比べると自然な返戻の判断は難しい。雑誌や大部な報告書など多くの構成要素を含む巨大なツリー構造も相当あり、あまりに遠く離れた組合せをヒットと見なすとノイズが増え、また行き過ぎた集約を加えると結局最上位の資料タイトルレベルが返戻されることが増えて、なるべくピンポイントに返戻するという方針から外れていく恐れがある。そのため一定の制限距離（親等）を設定してその範囲内のものをヒットと見なすこととしているが、適切な制限距離を判断するデータ分析は不足しており、今後の課題である。また前節で述べたように、資料による構成要素間の関連性の強さの差異に対する考慮が求められるが、この点は実装にいたっていない。

依田らの指摘にある返戻方式の問題点については、従来の「トップダウン方式」を原則としながらも「ボトムアップ方式」を一部導入した。すなわち、下位メタデータの多数がヒットした場合には集約して上位メタデータを返戻する。ただし、多数の記事が並列に並ぶ雑誌や報告書でこれが不用意に行われるとピンポイントアクセスが阻害される。防止するには前節に述べたように下位メタデータ中での「ヒット率」を判断材料にすべきであるが、比率算出の母数である下位メタデータ総数をその都度得る処理のシステム負荷といった問題があり、「ヒット数」で判断する実装となっている。弊害防止のため比較的多い値を分岐点としており、柔軟でわかりやすい返戻という効果は十分に実現しているとはいえない。

5. おわりに

以上、震災デジタルアーカイブのメタデータ作

成・蓄積・検索について、資料中の構成要素と資料構造の取扱いに絞って述べてきた。(1) 資料中の構成要素について重層的にメタデータ記述を行うこと、(2) 各構成要素を独立した操作対象とするとともに、資料構造を表現する階層構造情報を管理すること、(3) 論理積検索において直列的な階層関係にあるメタデータ群を一体のものにとらえる仕組みを持つこと、(4) さらに状況に応じて資料中の共出現をとらえるなど拡張的な論理積検索の機能を持つこと、が主要な特色である。なお、当然ながらメタデータには、各メタデータセットの意味論的問題、構文的問題、あるいは主題情報とその検索のあり方など、他にも多くの問題があり、震災デジタルアーカイブの特徴点もあるが、それらはいずれ稿をあらためて論じたい。

4 章で述べたように、現行システムで対処できていない問題もいくつかある。また客観的に見て、改善の最大の眼目である4種の論理積演算の実装は机上の論理に基づくもので、当否・有効性は証明されたとはいえない。今後、複雑な演算をわかりやすく説明するヘルプ機能を充実させるとともに、アクセスログ解析等によって利用者の検索行動を分析していくことが必要である。

歴史的資料が中心で1館あたりのデジタル化資料数も少なかった時期には、和装本・文書・巻物・絵図といった多様な資料の一つ一つをどのように画面上で再生するかに最大の精力が注がれるのもやむをえないが、次第に規模が大きくなり、また現代資料のデジタル化に本格的に取り組むようになれば、情報組織化の問題に直面することは避けられない。「はじめに」でも述べたが、この点に関する図書館関係者の意識はいささか稀薄であるように思われる。全文検索エンジン等のシステムの仕組みや、メタデータにおけるダブリン・コアに代表される「標準」を所与の条件ととらえて対応していく方向性が一般的で、「当事者意識」があまり感じられない。まずシステムありき、ではなく、対象となる資料の実相をとらえて適切にモデル化し、利用者の情報要求を想定して蓄積・検索手法を考えていくことを、主導的に行っていかななくてはならない。

本稿で述べた震災デジタルアーカイブの手法は一定の応用性を持つと考えるが、出発点は2章で述べた震災資料の特色(著しい「多様性」)である。多くの電子図書館でそれぞれの対象資料群の特性をとら

えた情報組織化の試みがなされることで、より汎用性を持った知見が蓄積されていくのであり、そうした実践が求められている。

謝辞

本稿は、大阪教育大学教育学研究科健康科学専攻(夜間大学院)の修士論文(論題:「地域資料のデジタル化と情報の蓄積・検索方式—震災デジタルアーカイブからの考察—」)の一部をもとに加筆等を行ったものです。塩見昇先生をはじめご指導いただいた先生方に深く感謝いたします。

- 1 国立国会図書館「近代デジタルライブラリー」
<<http://kindai.ndl.go.jp/>>. [引用日:2004-02-09]
- 2 渡邊隆弘「図書館目録とメタデータ」『図書館界』53(2), 2001. 7. pp. 126-133.
- 3 神戸大学附属図書館「震災文庫」
<<http://www.lib.kobe-u.ac.jp/eqb/>>
[引用日:2004-02-09]
現在本 WWW サイトのトップページは「震災文庫」と表記しているが、神戸大学電子図書館システムではコンテンツを総称して「デジタルアーカイブ」と呼称しているところから、本稿では「震災デジタルアーカイブ」と呼ぶこととする。
- 4 古川肇「構成部分の記述: 将来の目録」『整理技術研究集録』1, 1993.3. pp. 5-12.
- 5 Larson, Ray R. "Between Scylla and Charybdis: Subject Searching in the Online Catalog" *Advances in Librarianship*, 15, 1991. pp. 206-207.
このレビューによると、目次情報の利用は1970年代後半に既に試みられている。
- 6 永田治樹『学術情報と図書館』丸善, 1997.5. pp. 119-121.
- 7 *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev. Chicago: ALA, 2002.
「分出」の章はAACR2 初版(1978)からある。
- 8 『日本目録規則 1987年版 改訂2版』日本図書館協会, 2001.
「構成書誌単位(構成単位)」は、「固有のタイトルを有しているが、形態的に独立していない、資料の構成部分を記述対象とする」書誌単位と定義されている(p361 用語解説)。
なお、この定義は1987年版初版(1987)から変更されていない。
- 9 IFLA. *Guidelines for the Application of the ISBDs to the Description of Component Parts*.

- Washington, D.C.: Library of Congress, 2003.
(Reproduction with corrections of the 1988 edition)
<http://www.ifla.org/VII/s13/pubs/Component_Parts_final.pdf>.
[引用日:2004-02-05] 本ガイドラインには、古川の解説論文がある。
古川肇「「ISBDを構成部分の記述に適用するためのガイドライン」の紹介」『整理技術研究』27, 1999. pp. 1-10.
- ¹⁰ 東京大学の「ブックコンテンツ」、国立情報学研究所の「WebCat Plus」が目次情報を含んだ大規模な書誌データベースである。
東京大学情報基盤センター図書館電子化部門「ブックコンテンツ・データベース」c1999
<<http://contents.lib.u-tokyo.ac.jp/contents/top.html>> [引用日:2004-02-09]
国立情報学研究所「Webcat Plus」c2002
<<http://webcatplus.nii.ac.jp>>
[引用日:2004-02-09]
- ¹¹ ISBD(CP)ではさらに下位部分を持つ「多段階構成部分 (multi-part component)」の存在が意識されているが、その表現方法が一定しておらず「検討が不徹底であるように見える」ことを古川は指摘している (注4前掲論文)。
- ¹² 同様の論旨を、注2前掲論文等で発表している。これに対し古川は「一般目録についての主張」として、「著者は、自らの著作を何よりも総体として理解され…ることを望んでいるはず」なので著作の「一部分を前後の文脈から引き離して検索対象」とするのは好ましくないと述べている。
古川肇「目録の構造に関する試論」『資料組織化研究』44, 2001. 7. pp. 1-10.
筆者は、著作者の意図によるまとまりを基本的単位としながらも、利用者の必要に応じてより細かな情報単位へのアクセスを可能にすべきだと考えている。もちろん、断片を切り取った提示によって著作者の意図が誤解される危険は常に存在しており、防止するためには一方で著作・資料の全体像を的確に伝える情報提供が必要と考える。
- ¹³ 長尾真「目次情報などを利用した図書・文献検索方式」『情報の科学と技術』42(8), 1992. pp. 711-718
石川徹也ほか「電子図書館 Ariadne の開発 (3) 検索支援機能」『情報管理』38(6), 1995. pp. 538-548
- ¹⁴ 震災文庫については、筆者によるものを含め、次

の報告がある。

- 稲葉洋子「阪神・淡路大震災関連資料を収集して：灰色文献とインターネット」
『専門図書館』171, 1998. 9. pp. 81-85.
稲葉洋子「震災資料の保存と公開：神戸大学「震災文庫」を中心として」
『大学図書館研究』55, 1999. 3. pp. 54-64.
渡邊隆弘「『震災文庫』のこれまでとこれから—電子図書館を中心に」
『Academic Resource Guide』055, 2000. 2. 15.
<<http://www.ne.jp/asahi/coffee/house/ARG/compass-029.html>> [引用日:2004-02-09]
渡邊隆弘「『震災文庫』サイトのそれから—一次情報のデジタル化を中心に」
『Academic Resource Guide』096, 2001. 4. 15.
<<http://www.ne.jp/asahi/coffee/house/ARG/compass-043.html>> [引用日:2004-02-09]
- ¹⁵ ただし、博物資料にあたるような「遺物」までは対象とはしていない。また、公開を目的とする収集であるからプライバシー等の面で秘匿性の高い種類のもの(避難所の業務日誌や行政の部内秘文書など)はあまり所蔵していない。
- ¹⁶ その時点では、PC上のデータベースソフトで作成した情報をWWWサーバに転送し、職員の自作プログラムで検索させるという、手製のシステムであった。このシステムは1999年7月まで、4年近くにわたって安定して使われ続け、開設時点で約1,800件だった収集件数は、1999年7月には約17,000件に達している。
- ¹⁷ ただし、この時点では目録データの補完に主眼があり、ネットワーク上での資料閲覧に耐えうる画像解像度ではなかった。
- ¹⁸ 「電子図書館化推進経費」は特定の国立大学に集中的に予算を投下することで、「先導的」な電子図書館開発や内外に広く利用できるコンテンツの充実をはかったもので、奈良先端科学技術大学院大学・京都大学・筑波大学・東京工業大学に次いで5番目の導入(図書館情報大学にも同時に導入された)であった。なお、神戸大学電子図書館システムの全体的事項については、下記の文献を参照されたい。
渡邊隆弘「神戸大学電子図書館システムにおける「電子アーカイブ」の構築」
『デジタル図書館』16, 1999. 11
<http://www.dl.ulis.ac.jp/DLjournal/No_16/1-watanabe/1-watanabe.html>
[引用日:2004-02-09]

19 震災デジタルアーカイブにおけるメタデータについて、次の論文を発表している。

渡邊隆弘「震災アーカイブにおけるメタデータ的设计」『人文科学とコンピュータシンポジウム論文集』, 情報処理学会, 2000. 12. pp. 89-96.

20 本アーカイブでは全メタデータに共通なデータ項目設定ではなく、リソース種別に応じて必要項目が設定される。また、画面上のメタデータ表示（項目ラベル等）もリソース種別に応じて変化する。本稿ではこれ以上ふれないが、注 19 前掲論文を参照されたい。

21 パス情報の先頭部分はツリーごとに付与された特定の番号が埋め込まれており、同一ツリーに属するメタデータの抽出が可能であるが、実際には高速検索のため各メタデータレコードに属するツリーの「最上位メタデータ ID」を別途格納して抽出に用いていた。

22 XML (eXtensible Markup Language) 等のマークアップ言語とメタデータの関わりについては注 2 前掲論文で述べたが、対象データの各要素を<NODE>〜</NODE>といった書式の「タグ」で囲むことにより、文書・データの構造を明確に表現することが可能である。複雑な構造に対応できるほか、タグ設定は対象データにあわせて設計者が DTD (Document Type Definition 文書型定義) として定義するため、どのようなデータにも対応できる柔軟性にも優れている。

23 抽象化された図ではいささかわかりにくいかもしれないが、「章」「節」「項」といった具体を想定すると、レコード 6 (例えば章) とは下位の構成要素（節や項）を包含した内容に相当するものであり、対応する<S>〜</S>を下位要素を囲む形で設定することが首肯されよう。

24 主題索引法における「アップワードポスティング」の考え方と同じことである。

25 1.で述べたように、“Ariadne” (注 13) では階層上下の出現順序を考慮して検索精度の向上をめざしているが、本アーカイブでは出現順序を考慮していない。Ariadne は学術的な内容の図書を主対象としているが、震災資料は内容・形式面でより幅広い内容を含んでいるので、論理的に構成された資料構造を前提とした細かな処理は行っていない。

26 依田平、大月一弘、森下淳也、清光英成「デジタルアーカイブに対する効率的な検索の提案：神戸大学電子図書館システムを例として」『人文科学とコンピュータシンポジウム論文集』情報

処理学会, 2001.12. pp. 259-266.

27 コピー入力による運用には、データ作成時の手間は許容範囲であっても、いったん作られたデータの修正に多大な労力を要するという問題もある。

28 現行システムについては、次の論文を発表している。

依田平、渡邊隆弘、大月一弘、鳩野逸生、岩杉大輔「多様な資料構造に対応したデジタルアーカイブシステム —神戸大学電子図書館アーカイブ検索システム—」『情報処理学会研究報告』2003-FI-73, 2003.11. pp. 45-52

29 4つの論理積演算子について、プルダウンメニューから選択できる画面もある。

30 入力されたキーワードごとに検索結果集合をいったん作成し、結果リスト間の比較を行う。キーワードごとの結果集合が大きいとシステム負荷が大変高く、現状では非常に多数のデータがヒットする場合には、本来指定された論理積演算をあきらめ、1 資料内の共出現にのみ着目して最上位データを返戻する仕様となっている。