

複合的メタデータによる文化情報資源へのアクセス可能性の提示

矢代 寿寛¹ 大山 敬三²

¹ 総合研究大学院大学複合科学研究科情報学専攻

² 国立情報学研究所

〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: { yashiro , oyama } @nii.ac.jp

抄録

デジタルアーカイブの増加や書籍のデジタル化に伴い、文化情報資源のメタデータと資源形態が多様化したため、エンドユーザだけでなく管理者側も網羅的にアクセスパスを把握することが困難となっている。本稿では、貴重書をはじめとする文化情報資源にもたらされたアクセスの問題を解決するため、文化情報資源のアクセス可能性を通覧的に提示する手法を提案する。既存メタデータに基づくアクセス可能性を通覧的に提示することで、エンドユーザのアクセス支援と管理者側のコレクションマネジメントとリソースマネジメント支援を可能にする。手法の持続可能な実装のあり方として、今後検討していく二通りのシステム設計モデルについて述べる。

Presentation of accessible cultural resources by multiple metadata

Kazunori YASHIRO¹ Keizo OYAMA²

¹ School of Multidisciplinary Science, Department of Informatics, Graduate University for Advanced Studies

² National Institute of Informatics

2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: { yashiro , oyama } @nii.ac.jp

Abstract

With the increase of digital archives and digitize books projects, digital objects providers also need to comprehend access paths as well as end-users. This paper describes a presentation of accessible culture resources take rare book an example. We discuss a method of overview presentation of accessible cultural resources by multiple existing metadata. In addition, we argue that the method is useful for end-users and digital objects providers. The last part of this paper will focus on an implementation of system models.

1. はじめに

近年、大学等において期限付きのプロジェクトや記念事業といった形で、中小規模デジタルアー

カイブが多数構築されている。また、大手検索エンジン事業者、大学図書館、コンソーシアムらの協働による書籍デジタル化プロジェクトが多数行われている[1][2][3]。このため、かつては博物館(Museum)、図書館(Library)、文書館(Archives)と

いう情報資源の保存や提供を担う社会教育施設(以下、MLA)で、現物資料を参照するほかになかった文化情報資源であっても、高精細画像やテキストデータ等の資源形態からエンドユーザが選択的に利用できるようになった。他方で、文化情報資源の管理者はデジタルオブジェクトのプロバイダとして、現物資料以外の情報管理や権利処理に対応する必要が生じてきた。

著作権や資料保全の関係上、デジタル化される文化情報資源は、博物館資料とも図書館資料ともとりうる貴重書や古地図、歴史資料が多くみられる。過去に行政文書としても扱われた場合は文書館資料でもあるため、管理者の組織的境界はさらに曖昧である。このような MLA 横断的な文化情報資源のデジタル化作業は、管理者やプロバイダごとに行われており、主に実施組織が属する分野の標準的なデータ形式やメタデータ記述が利用されている。営利企業であれば自社・自陣営規格が利用されることとなる。

図書館または文書館でのメタデータ記述では、ある程度の国際標準化がなされているため横断検索やデータ共有も可能といえる。しかし、博物館のメタデータ記述では、資料の多様さや同定識別の困難さから国際的な標準化が容易ではなく、ガイドラインの制定やモデルの提案に留まっている[4][5]。そのため、同じ文化情報資源に対して異なる組織がデジタル化を行っている、アクセシビリティやファインダビリティが異なるデジタルオブジェクトが同時に分散して存在することになる。ここに現物資料が加わる文化情報資源へのアクセスパスを網羅的に把握することは、エンドユーザだけでなく管理者側でも困難といえる。

この状況は、学術情報における「適切コピー」の問題とみなすことができるように思われる。しかし、いくつかの点で相違がみられるため、個々の現物資料やメタデータ記述を含んだ範囲での対応が必要と考える。また通常、管理者やプロバイダに位置づけられる MLA の多様さと行財政的環境などを考慮すると、エンドユーザにとってのアクセシ

ビリティ向上だけでなく、MLA がデジタル化作業や現物資料を効率的に管理・運用できるような支援が必要となってくる。

そこで本研究では、エンドユーザのアクセス支援および MLA のコレクションマネジメントとリソースマネジメント支援を目的とした、文化情報資源へのアクセス可能性を通覧的に把握できるような仕組みの提案と構築を行う。これにより、文化情報資源への持続可能なアクセス確保を目指す。

本稿ではまず、MLA 横断的な文化情報資源である貴重書を例に、メタデータとアクセスパスが多様化している現状について整理する。その上で、現在の文化情報資源におけるアクセス可能性の問題について、従来の「適切コピー」問題と異なる点を考察する。これを踏まえ、既存メタデータを複合的に応用した文化情報資源への通覧的なアクセス可能性提示を提案する。最後に、提案手法の実装に今後必要となる分散型と集約型のシステム設計モデルについて、それぞれの特性や利点、欠点を考察する。

2. 文化情報資源の多様化

多様化した文化情報資源のメタデータとアクセスパスについて、貴重書として扱われる福澤諭吉著「学問のすすめ」を例に述べる。「学問のすすめ」は、1872年の初篇から1876年の第17篇まで分冊刊行後、1880年に合本で再出版された。ここでは、初篇とそれを含む合本の初版および第2版に絞り、デジタルオブジェクト、現物資料、復刻版その他の順で挙げる。

把握することができた2008年2月の時点で利用可能な「学問のすすめ」のデジタルオブジェクトの、提供者、元の形態と版、メタデータ項目数、メタデータ標準への対応、提供データ形式、色表示、表記ゆれへの対応、内容テキスト検索の可否、現物資料の有無、現物資料へのアクセス可能期間、備考を表1に示す。

表について補足する。慶應義塾図書館所蔵の

合本初版と合本再版をデジタル化した Google Book Search の場合、「學問すゝめ：自第一篇至第十七篇」をタイトルとしているため、「学問のすゝめ」等の異表記では検索できない。また著者名も「福沢」ではなく「福澤」でなければならない。メタデータの記述項目は「タイトル」「著者」「発行者」「発行年」「ページ数」である。次に、慶應義塾図書館デジタルギャラリーでは初篇の初版と 17 分冊本の第 2 版を公開しているが、「学問のすゝめ、初編」とあり、初篇のメタデータ記述とわかる。記述項目は「タイトル」「別タイトル」「著者」「出版地」「出版年」「識別番号」だが、検索は内容テキスト全文である。そのため「別タイトル」の「Encouragement of learning」等では検索できない。合本初版をデジタル化した国立国会図書館近代デジタルライブラリーでは、通常の OPAC と同様の書誌が作成されている。筑波大学図書館の場合は貴重書としてではなく、通常の OPAC の検索結果にファイルがリンクされている。京都大学図書館では、貴重書専用の検索システムを利用するが、書誌事項は通常の図書より簡略化されているとみられる。なお、国文学研究資料館のデータベースは現時点では利用できないため、詳細は省く。

表 1 デジタル化された「学問のすゝめ」

提供者	Google	慶應義塾図書館	国文学研究資料館	国立国会図書館	筑波大学図書館	京都大学図書館
形態と版表示	合本	分冊	分冊/合本	合本	分冊	分冊
メタデータ項目数	5	7	不詳	21	17	9
メタデータ標準	不明	不明	不明	NCR	不明	不明
データ形式	jpg, text, pdf, html	jpg	jpg, jpeg2000	jpg, jpeg2000	pdf	jpg
色表示	カラー/モノクロ	カラー	カラー	モノクロ	モノクロ	カラー
表記ゆれ対応	なし	あり	なし	あり	あり	なし
内容テキスト検索	あり	あり	なし	なし	なし	なし
現物	なし	あり	あり	あり	あり	あり
現物へのアクセス	なし	平日8:45-17:00	平日9:00-18:00	平日9:30-17:00	平日9:00-12:15, 13:00-17:00	平日9:00-11:45, 13:00-16:45
備考		要予約	移転のための閉室中	要申請 マイクロフィルム	要申請	要予約

デジタルオブジェクトの次に、現物資料のみを

所蔵・提供している組織を挙げる。国立情報学研究所の Webcat Plus によると、国立国語研究所および国公立大学 10 校以上が初篇を含む分冊本を所蔵していることがわかる。国文学研究資料館の近代書誌・近代画像データベースでは、神戸大学と酒田市立光丘文庫を所蔵者とする書誌があり、所蔵が推定できる。国立国会図書館の総合目録ネットワークシステムでは東京都立図書館等公共図書館での所蔵が確認できる。この他にも個人蔵の可能性などがあるが、現時点では確認する術を持たない。エンドユーザのコンテキストによっては、明治期出版の現物資料やそのデジタルオブジェクトが必要ではないといえる。復刻版や現代語訳、影印本が多数出版されているほか、青空文庫等が Web 上で公開しているテキストデータも利用できる。

以上は一貴重書の事例であり、文化情報資源全般の多様性を示してはいる訳ではない。しかし、公文書でもあった古地図や、時事風刺を含んだ歴史資料でもある浮世絵など、同様の状況は幅広くみられる。近代の文学や美術作品の著作権が消滅するに伴い、この傾向は加速すると考えられる。

3. 「適切コピー」問題と関連研究

事例で述べた文化情報資源のメタデータとアクセスパスの多様化は、ある種の「適切コピー」問題とみなすことができるが、いくつかの点で異なっている。解決すべき問題を適切に設定するため、まず従来の「適切コピー」について整理する。

「適切コピー」とは、学術情報の電子化によって図書館、アグリゲータ、学術出版者など複数の場所に同一内容の論文やテキストが存在する中で、ある情報資源からのリンク先として質やコストがユーザコンテキストに適したオブジェクトやサービスを指す。この「適切コピー」へどのようにユーザを導くかが従来の「適切コピー」問題である。「適切コピー」問題を構造的に把握するため、

FRBR[6]化した例を図 1 に示す。

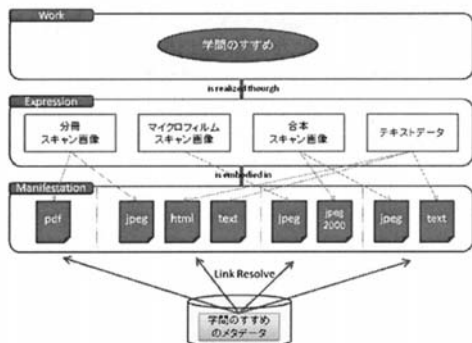


図 1 電子ブック・テキストの「適切コピー」例

この「適切コピー」問題を解決する方法としては、Caplaら[7]や Beit-Arie, Sompelら[8]のリンキング研究に基づいて開発されたリンクリゾルバを利用することが多い。CrossRef[9]の援用する DOI[10]と OpenURL[11]の規格を用いた商用の S・F・X[12]が代表的であり、大学図書館などが OPAC に導入している。

「適切コピー」の前提となる著作全体の構造把握には FRBR モデルが用いられる。FRBR については IFLA[13]が網羅的に関連研究を収集しており、また鶴田[14]のような電子情報資源に特化した独自モデルもみられる。両角ら[15]はリンクリゾルブのための FRBR に基づくメタデータスキーマモデルを提案している。

4. 文化情報資源のアクセスにおける問題

現在の文化情報資源にもたらされている状況は、従来の「適切コピー」問題と大きく異なっていると考えられる。異なる点について、大きく分けて情報資源の差異とプロバイダの差異の 2 つが挙げられる。

まず第 1 に、文化情報資源の性質が均一で不変ではない。「適切コピー」となりうる情報資源は

学術論文や図書資料であり、これらは資源形態が変化しても基本的に内容は変化しない均一的な性質といえる。しかし、文化情報資源は資源形態ごとに固有の性質を持つものがみられ、形態変化によって情報が失われたり、新たな情報が付け加えられたりする。例えば複製本のような複製品は、オリジナルの固有の情報がいづらか失われるが、複製固有の情報も持つことになる。出版事項などはこれにあたる情報である。また、性質が異なるということは要求と利用のされ方が異なるということでもある。文化情報資源は、特に博物館や文書館の資料の場合、単体よりもメタデータや解説、来歴といった二次情報と合わせた形で一般に要求、利用される。そのため、何がどのような組み合わせであればエンドユーザにとって適切であるかを機械的に判断することは困難であり、コンテキストに依存することとなる。例えば、デジタル画像は高精細だが解説が欠けていたり、メタデータが更新されていないということがあれば、別途異なったコンテキストでの資源形態やメタデータが適切となりうると考えられる。この文化情報資源の性質を考慮した場合の、「適切コピー」としてとりうる形態を FRBR モデルに近づけた形で構造的に図 2 に示す。

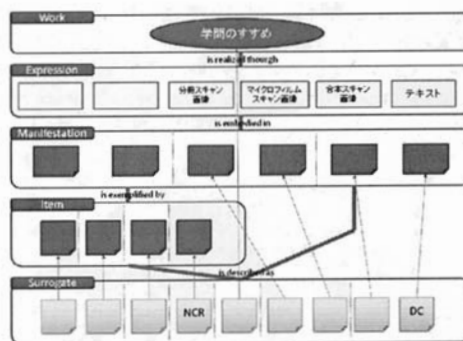


図 2 文化情報資源の擬似的 FRBR 化例

第 2 に、文化情報資源の主要なプロバイダとなる MLA それぞれの役割が同一ではない。MLA

を単なる管理者としてみた場合には、収集、飼育・保管、調査・研究、展示・提供という 4 つの主たるミッションを、それぞれの割合は異なるが共有している。しかし、プロバイダとしてみるとミッションにも違いが生じている。

図書館には、非営利無料の原則が国際的に適用されており、知る権利の保障、学習支援、セーフティネットといった多様な役割を持つ。また、文書館も図書館と同様に非営利無料であり、知る権利の保障とともに行政の説明責任を果たす場でもある。ただし一般に、学習目的で利用されることは少ない。そして博物館は、国際的には原則無料とされているが私立博物館も数多く、また公立であっても例外的措置として入場料を徴収するなどしている。このため、仮に同一内容の文化情報資源の現物資料が他組織でも展示され、一般の利用に供されていた場合は、それが適切であろうとなかろうと、自館へエンドユーザを導くインセンティブが働くことになる。「適切コピー」においては、同一的な内容のリソースを持つことになる図書館、学術出版者、アグリゲータらプロバイダは、どこが利用されようと、利用時の料金は最終的に著作権者へ行きつくため、広告収入などを除けば自サイトへ導くインセンティブに乏しいといえる。そもそもユーザはどのプロバイダであろうと利用できれば構わないという認識がリンクリゾルバによるナビゲーションを可能にしていると思われる。

これらの違いを踏まえると、この状況を「適切コピー」とみなしたままでリンクングによる解決策を検討すべきでないと考えられる。そこで本研究では、エンドユーザのコンテキストに対応したアクセス支援に留まらないこの問題を解決するため、文化情報資源への通覧的なアクセス可能性提示を提案する。

5. 文化情報資源へのアクセス可能性提示

文化情報資源について、何時アクセス可能か、どの程度のコストでアクセス可能かといった情報を、

アクセス可能性として通覧的な提示を行う。文化情報資源はそのアクセス可能性を決める要因として、所在、利用状況、プロバイダの利用可能期間、プロバイダの利用コストの 4 つが挙げられる。長期的には記録メディアの寿命が影響する資源形態もあるが、標準等の決定がないためここでは考慮しない。

所在は、文化情報資源の現物資料を所蔵または提供している組織内における所在である。例図書館では、開架書庫、閉架書庫、貴重書室等がこれにあたり、目録でも用いられている項目である。ただ、この所在は所有者の組織内であるとは限らず、貸出や輸送によって他組織内が所在となることもある。このとき、通常は次に述べる利用状況が変化するのみで所在は変化しないが、本手法ではこれを常に実際の所在と対応させる。

利用状況とは、MLA にほぼ共通する管理情報で、貸出中、修復中、デジタル化作業中などがある。本手法では、所在の情報よりも利用状況を優位にしてアクセス可能性を決定づける。

プロバイダの利用可能期間とは、現物資料を提供する MLA においては開館日時、デジタルオブジェクトを提供するプロバイダにおいてはメンテナンスや障害が起きていない日時である。本手法では、これをアクセス可能性を決定する最大の要因として利用する。

プロバイダの利用コストとは、利用料という直接コストと、プロバイダに至るまで支払う間接コストがある。本手法では、ユーザの属性情報である年齢層と大まかな所在地を選択的に利用することで、両者を算出する。これはアクセス可能性には直接影響しない要因として扱う。

本稿では、所在をロケーションメタデータ、利用状況をステータスメタデータ、プロバイダの利用可能期間と利用コストを併せてプロバイダメタデータとする。これらを組み合わせることで文化情報資源のアクセス可能性とする。これを通覧的に提示することで、何が何時アクセス可能または不可能となるかが容易に把握できる。これにより、エンドユーザへ

のアクセス支援だけでなく、プロバイダ側へのマネジメント支援を行うことができる。例として、デジタルオブジェクトが既に存在する書籍のデジタル化優先順位を変更する、現物資料へのアクセスの空白期間を利用して普段は収蔵庫にある類似資料を展示する、などが考えられる。このアクセス可能性からマネジメントを行い、マネジメントによりアクセス可能性が変わっていくことでサイクルが生じると考えられる。このサイクルによって、リアルアーカイブとバーチャルアーカイブが相互に補完され、文化情報資源への持続可能なアクセスが確保されると期待する。

6. システムの設計と2つのモデル

6.1. システムの機能とユーザインタフェース

手法を実現するシステムについて、機能とユーザインタフェースを中心に述べる。本システムでは4つの機能と、3種類のユーザインタフェースによる操作を想定している。

1つ目は、文化情報資源の現物・非現物を問わずに各形態を検索する機能である。この結果から、ロケーションメタデータ、ステータスメタデータ、プロバイダメタデータを用いて、アクセス可能性を3段階に分けた形で提示するのが2つ目の機能である。現時点で設定する3つの段階は、1)今現在アクセス可能であり、1時間後も利用可能である、2)今現在アクセス可能であるが1時間以内に利用不可能になる 3)今現在利用不可能である、としている。2つの機能を実現するインタフェースを図3に示す。

インタフェースの基本画面では、文化情報資源の各形態を、期間、地理、コストによりソート可能な形で通覧的に表示する。各形態のレコードは、可能であれば各文化情報資源へのリンクとするが、基本的にはアクセスのナビゲーションとサロゲートとして提供する。レコードに `permalink` を付与する

ことで、特定形態のみをブックマークしたり、モバイル端末で利用することを可能にする。このインタフェースの特徴として、エンドユーザ用とプロバイダ用が表裏一体となった形をしている。すなわち、エンドユーザ側でアクセス可能として提示されている文化情報資源は、プロバイダ側ではアクセス不可能と提示される。基本的に、プロバイダがアクセス可能な状態というのはエンドユーザがアクセスできない状態であるといえる。例外としては輸送などがありうるが、この場合はどちらからもアクセスできない状態として提示する。

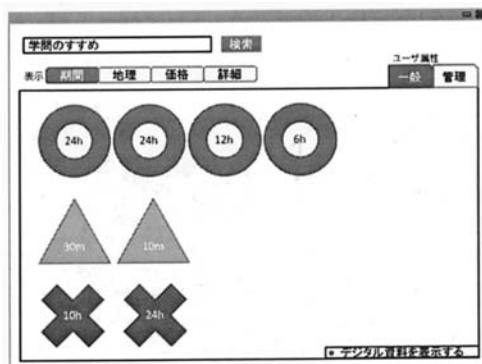


図3 基本的なユーザインタフェース

残りの2つの機能は、プロバイダ側だけに提供され、それぞれが専用の操作インタフェースを持つ。1つは、アクセス可能状況の変更機能である。何らかの理由により、通常のアクセスサイクルとは別に文化情報資源をプロバイダ側が利用する必要が生じた際、一時的にアクセス不可能状態へと移行させることができる。この場合、変更者による記述でアクセス不可能理由とアクセス者がプロバイダ側で把握できるようにする。もう1つは、アクセス可能なデジタル資料における1層深いレイヤでのバリエーションの通覧機能である。MLAの業務で撮影・作成されるデジタル画像は、エンドユーザに示されるよりも数多くなる。例えば、彫像や塑像などでは正面以外にも左右斜側面や背面等様々な角度からの画像がある。これは企画展示な

どに際してその都度撮影がなされたり、画像の劣化による再撮影などが原因のほか、学芸員らによる調査研究や不均質な管理専用画像の撮影が行われるためである。そのため、エンドユーザには不要と思われる画像だが、情報管理の都合上、プロバイダ側には提示する。以上が、システムで提供する予定の機能である。

6.2 システムモデルの検討

システムを実装するための設計モデルとして、大きく分けて分散型と集約型の 2 つが考えられる。それぞれの構成を図 4 に示す。

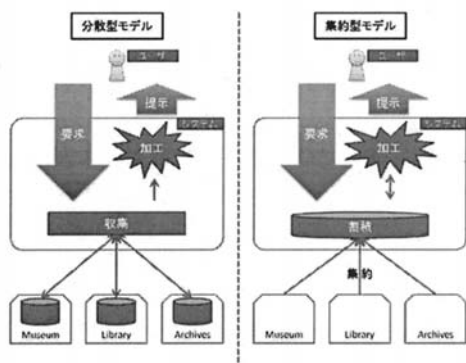


図 4 システムの設計モデル

両者の特性、実装で得られる機能および効果の質的な違い、そして現在の日本の MLA が置かれている環境での実現と持続可能性について今後検討すべき事項を述べる。

まず図 4 の左側に示した分散型では、各文化情報資源のロケーションメタデータとステータスメタデータが日常的に作成、管理されていることが前提となる。これらを横断的に検索することで、提示を要求される都度収集し、加工する。利点として、分散環境一般にいえる負荷とリスクの分散がある。欠点としては、各組織が共通のプロトコルやデータ形式に対応しなければ実行が困難となること

ある。また、各組織へ個別にアクセス可能かどうかを訪ねると実質的には同義であり、手間の軽減目的以外でシステム化する意義に乏しいと考える。

次に、図 4 の右側に示した集約型では、予めプロバイダ側からのデータ提出やメタデータハーベスティングによって収集・加工しておき、提示するものである。利点として、デッドロックやトランザクション処理への対応が可能である。欠点として、集約による負荷とリスクの集中のほか、アクセス可能性の変化に即時に対応できないことがある。

どちらもプロバイダの持つロケーションメタデータ、ステータスメタデータ、プロバイダメタデータの 3 種を利用するが、メタデータを収集するタイミングが異なる。そのため、メタデータの加工や提示における付随的な要素に違いが生じる。

実現と持続の可能性の面からみると、分散型は比較的情報化の遅れが見える博物館において、有用な統合システムの普及がなされれば、本システムの負担は軽減され、MLA 側も独自に発展させていくことが期待される。一方、集約型でのメタデータ利用は文化庁の文化遺産オンライン等で既に行われているため、現在の日本の MLA 環境での実現は可能と考えられる。しかし、提示機能の高次化等は本システム側で行わなければならない、開発の負担となることが危惧される。両者の利点、欠点について議論を深め、システムの機能や効果の質的な面に及ぼす影響について、ハイブリット型を含めた検討が必要と考える。

7. おわりに

本稿では、文化情報資源への持続可能なアクセスを目指し、MLA 横断的な文化情報資源へのアクセス可能性の通覧的な提示を提案した。アクセス可能性の提示には、ロケーションメタデータ、ステータスメタデータ、プロバイダメタデータの 3 種を複合的に用いる。日常的な業務に組み込まれた各既存のメタデータを用いることで、MLA 側の負担を軽減することができる。このアクセ

ス可能性の提示により、エンドユーザにとってのアクセシビリティが向上するだけでなく、プロバイダ側のリソースマネジメントやコンテンツマネジメントも支援することができる。このマネジメントとアクセス支援のサイクルを実現するシステム的设计モデルについて、分散型と集約型、それぞれの利点と欠点を考察した。

今後は、現在の日本の MLA 環境でも実現と持続可能なシステムの実装に向けて、モデルの詳細な検討を行う。システム構築後は、MLA のスタッフやエンドユーザによるアウトプットとアウトカムの評価を予定している。

参 考 文 献

- [1] Google. Google Book Search .
<<http://books.google.com/>>. (Accessed 2008-03-01).
- [2] Microsoft. Live Search Books .
<<http://search.live.com/results.aspx?q=&scope=books>>. (Accessed 2008-03-01).
- [3] Open Content Alliance.
<<http://www.opencontentalliance.org/>>. (Accessed 2008-03-01).
- [4] ICOM. International Guidelines for Museum Object Information .
<<http://cidoc.mediahost.org/guidelines1995.pdf>>. (Accessed 2008-03-01).
- [5] 東京国立博物館. “ミュージアム資料情報構造化モデル”.
<<http://webarchives.tnm.jp/docs/informatics/smmo>
[i/](http://webarchives.tnm.jp/docs/informatics/smmo)>. (Accessed 2008-03-01).
- [6] Robert Maxwell. FRBR : a guide for the perplexed. Chicago, American Library Association, 2008, 151p, (ISBN 9780838909508).
- [7] Priscilla Caplan, William Arms. Reference Linking for Journal Articles . D-Lib Magazine. 1999, Vol.5, No.7/8, <
<http://www.dlib.org/dlib/july99/caplan/07caplan.html>>. (Accessed 2008-03-01).
- [8] Beit-Arie, O. et al. Linking to the Appropriate Copy: Report of a DOI-based prototype . D-Lib

Magazine.

<<http://www.dlib.org/dlib/september01/caplan/09caplan.html>>, (accessed 2008-03-01).

[9] PILA. crossref.org .
<<http://www.crossref.org/>>. (Accessed 2008-03-01).

[10] IDF. The Digital Object Identifier System .
<<http://doi.org/>>. (Accessed 2008-03-01).

[11] NISO. ANSI/NISO Z39.88 -2004 The OpenURL Framework for Context-Sensitive Services .

<http://www.niso.org/standards/standard_detail.cfm?std_id=783>. (Accessed 2008-03-01).

[12] Ex Libris Ltd. SFX: Overview .
<<http://www.exlibrisgroup.com/category/SFXOverview>>. (Accessed 2008-03-01).

[13] IFLA. FRBR Bibliography . <
http://www.ifla.org/VII/s13/wgfrbr/FRBR_bibliography.pdf >. (Accessed 2008-03-01).

[14] 鶴田拓哉. 電子資料を対象にした FRBR モデルの展開. 日本図書館情報学会誌. 2006, Vol.52, No.3, pp.173-187.

[15] 両角彩子, 杉本重雄. 利用者の特性と環境に応じたリソース選択のためのメタデータスキーマモデル. 情報処理学会研究報告.情報学基礎研究会報告. 2005, Vol.2005, No.108, pp.7-18.