

2 学術情報流通論

2.1 電子図書館の今後の展望

筑波大学大学院図書館情報メディア研究科教授・知的コミュニティ基盤研究センター
sugimoto@slis.tsukuba.ac.jp

杉本 重雄

1. はじめに

Digital Library¹(ここではデジタルライブラリと記す。)はインターネット上での重要な応用分野と位置付けられ、研究開発が積極的に進められてきた。こうした研究開発は大きく分けると新しい情報技術の開発を目的とする研究指向のものと、図書館等をベースにして実際的なデジタルライブラリ環境を提供するものに分けることができる。

前者の代表格は、すでに終了したものであるが、アメリカ National Science Foundation(米国科学財団, NSF) 他の共同助成で進められた Digital Library Initiative(DLI)であろう[1]。デジタルライブラリという分野から容易に理解できることであるが、そこには多様な技術要素が含まれており、またそれらを統合する技術も重要な要素となっている。そこで進められた研究は、応用分野と計算機科学・情報学分野を結びつけた新しい取り組みであり、かつ図書館に注目を集めたという点からは大きな役割を果たしたものであると言える。また、DLI からは数学・科学技術・工学分野での教育・学習資源の開発とそれらの利用のための基盤の研究開発を目指した National Science Digital Library(NSDL) プログが生まれている [2]。また、電子政府(Digital Government)に関する研究もデジタルライブラリに隣接する領域の研究分野であるといえる。

一方、後者には大学図書館や国立図書館などで進められた新しい図書館環境を提供するものや、大学等の組織による新しい情報資源の蓄積と流通の取り組みがある。アメリカ議会図書館の American Memory[3]に代表される歴史資料・貴重資料のデジタル化のプロジェクトは、図書館による典型的な取り組みと言える。大学図書館では、電子出版物の提供、資料の電子化と提供、インターネット情報資源へのアクセス支援、ネットワーク時代に適した新しい図

書館環境の提供といった幅広い取り組みがなされてきている。

出版、特に学術出版の環境がネットワークの発展によって大きな影響を受けた。現在では、大学図書館では電子ジャーナルは当たり前ものとなり、CrossRef のような電子ジャーナルならではのサービスも利用されている。こうしたサービスを支える識別子として出版社を中心に開発が進められた DOI(Digital Object Identifier)がある。

ネットワーク上での情報資源の流通や利用の進歩は、情報資源に関する情報の記述、すなわちメタデータの重要性を広く認識させることになった。たとえば、教育・学習資源、ビデオデータ、地理的・空間的情報、電子商取引、資料保存というように、多様な分野でメタデータの開発が進められている。こうしたいろいろなメタデータ規則の中で、インターネット上での情報資源の発見を目的として Dublin Core Metadata Element Set(DCMES)と呼ばれるメタデータ規則の開発が進められてきた。DCMES は、多様な分野の多様な情報資源に関する記述を前提として開発されてきたメタデータ規則であり、電子図書館分野では広く利用されている。DCMES の取り組みにおける基本的かつ重要な視点は、多様なコミュニティで作り出される多様な情報資源に対して相互利用可能なメタデータのための規則を上げることであった。そのため、エレメントなどの記述規則を決める際に、常に interoperability(相互運用性、相互利用性)の視点から議論が進められてきた。

こうした研究開発の中でデジタルライブラリにとって重要ないくつかの問題が注目されてきている。それらに電子資料、特に電子的に作られ流通する Born Digital と呼ばれる資料の長期保存の問題、デジタルライブラリ間やデジタルコンテンツの Interoperability の問題がある。

本稿では、これまでのデジタルライブラリの研究開発に関して概観する。デジタルライブラリは技術的視点からは、多様な情報技術を総合的に応用することが求められる分野である。ここでは、技術的話題についてはごく簡単に触れるにとどめる。図書館におけるデジタルライブラリ開発は研究図書館主

¹ 一般に用いられる訳は電子図書館あるいはデジタル図書館。本稿では、これらのことばを同一の意味で用いる。「図書館」ということばから、建物や施設を基盤としたサービスを連想させることが多いので、機能を中心に述べるため本稿ではデジタルライブラリと記している。

導で進んできた。ここでは図書館でのデジタルライブラリについて機能別にまとめている。メタデータはデジタルライブラリのみならずインターネット全般において重要な話題である。ここでは異なる種類のメタデータを紹介した後、メタデータの基本モデルについて論じる。最後にデジタルコンテンツの保存についても触れる。

2. デジタルライブラリ

2.1 デジタルライブラリのための新しい技術の研究開発の活動 歴史

1990年代にはいくつものデジタルライブラリあるいはそれに関連する領域の研究助成プログラムが欧米で進められた。前述の DLI はそうしたものの代表格である。DLI は 1994 年からの 4 年間に進められた Phase 1 (DLI-1) と 1998 年から始められた Phase 2 (DLI-2) に分かれている。DLI-1 では規模の大きな 6 プロジェクトが進められたのに対し、DLI-2 では大きなものから小さなものまで数多くのプロジェクトが進められてきている。ヨーロッパでは、イギリスの eLib プログラム、EU での第 5・6 フレームワークの中の Information Societies Technology (IST) プログラムの中で研究助成が進められた。こうした研究助成は情報技術の研究開発プロジェクトだけではなく、研究や高等教育のためのデジタルライブラリ環境の研究開発も進められた。

DLI が持っていた特徴は、複数の研究助成母体による共同助成であったことと、計算機科学、情報学の研究者、図書館や出版社といったコンテンツ所有者が参加するという異分野をつなぐプロジェクトであったことである。特に、DLI-2 の場合、助成母体が NSF、DARPA、NASA に加えて、議会図書館 (LoC)、人文基金 (NEH)、医学図書館 (NLM) に広がり、自然科学・技術分野から医学、人文社会科学分野までカバーした。DLI-1 が、計算機科学と情報学を基礎とするデジタルライブラリのための要素技術の研究を中心に進められたのに対し、DLI-2 はよりコンテンツ指向を強めたといえる。いずれの場合も、学際的色彩が強いことが特徴である。

情報技術分野におけるデジタルライブラリの国際会議の主要なものに、北米における JCDL (Joint Conference on Digital Libraries) とヨーロッパにおける ECDL (European Conference on Research and Advanced Technology for Digital Libraries) のシリーズがある。前者は ACM と IEEE-CS の情報検索、ハイパーテキスト、データベース等のコミュニティをベースにしたもので、後者は ERCIM をベースとするヨーロッパの情報技術の研究コミュニティを基盤としている。我が国でも、図書館情報大学・筑波大学での

国際シンポジウムや京都大学、名古屋大学での電子図書館国際会議などが開かれている。アジア地域をベースにしたものとしては ICADL (Int'l Conference on Asian Digital Libraries) が開かれてきている。

こうした会議は研究や開発の動向、デジタルライブラリの研究コミュニティを知る上で有用なリソースである。また、D-Lib magazine[4] や Ariadne[5] はデジタルライブラリ分野のオンラインジャーナルとして、プロジェクトや研究動向などを知る上で有用なリソースである。

2.2 図書館を基礎としたデジタルライブラリの開発

この 10 年間に図書館の電子的な情報環境は、大学図書館を中心として大きく変化してきた。現在では、多くの図書館がホームページを持ち、図書検索サービスなどを提供している。図書館におけるデジタルライブラリサービスの典型的な機能を下に並べる。

(1) 電子的に出版される資料の提供

電子的に出版される資料 (雑誌やデータベースなど) の提供は、出版物を利用者に提供するという意味では図書館としてごく基本的な機能である。学術雑誌の電子出版の発展とともに最近では数千タイトルの雑誌を提供する大学図書館も普通になってきている。文献[6]に国内の大学図書館の電子ジャーナルの利用状況が詳しく述べられている。

雑誌記事を電子化することによって複数の雑誌タイトルにまたがった仮想的なジャーナルのサービスや引用文献に関するサービスが可能になる。

一方、雑誌の電子的提供の背後にはいろいろな問題もある。学術雑誌のコストの増大の問題と寡占化といった問題が議論されている。小規模な学会などでは、独自に電子ジャーナルに対応することが必ずしも容易ではない。遡及的な電子化や電子ジャーナルの長期保存などの問題もある。

(2) 貴重資料や歴史資料を中心とする資料の電子化と提供

貴重資料の電子化は大学図書館を中心に広く行われている。最近では Google と有力な研究図書館との協調による大規模な電子化プログラムの開始が話題を呼んだ。

貴重資料や歴史資料の電子化は、図書館が持つ貴重資料そのものへのアクセスを制限する一方、電子的な複製物によりアクセス性を飛躍的に高めるといって、「保存とアクセス」の両方の観点から進められてきている。この機能は、図書館に蓄積された文化財をどこからでも閲覧できるようにするという視点から、重要な機能の一つであり、電子化はこれからも進め

られていくと思われる。一方、電子化された資料がどのように利用されるのか(あるいは利用できるのか)、実際に利用しやすくするにはどのようにすべきか、といった点についても十分に考慮されないといけない。単に、資料を電子化しインターネット上で提供するだけでなく、教育利用での利用性の促進や評価、新しい情報技術との組み合わせによる新しい分野の開拓といった活動が求められる。

(3) 大学などから出版される様々な情報資源の電子化とアクセスの支援

大学からは、学位論文やテクニカルペーパーなど通常の出版流通ルートには乗らない学術出版物が多く出版され、しかもネットワークを介して直接利用者に届けられている。大学などの組織が出版する情報資源の電子化や、電子化された資源へのアクセスを支援することは図書館にとっての重要な役割である。資料の電子化のために適切な技術を選ぶこと、適切なメタデータを与え組織化することなどが求められる。また、発信される情報の再利用や保存のための技術も求められる。そのため、図書館自身がこうした機能を実現するのみならず、発信者である組織に対して適切な技術支援をすることも求められる。大学や研究所といった組織に限らず、電子政府のような、巨大な情報発信元を想定してみるとこうした機能の必要性が理解しやすい。

このように、大学図書館にとって、大学が発信する学術情報資源を収集し、提供する機能(機関リポジトリ機能)は重要な機能である。欧米では様々な目的での学術リポジトリが開発され、それらの間の連携も進められている。かなり遅れをとっていると思えるが、リポジトリとその連携は重要な課題であると考えられる。

(4) インターネット上の情報資源へのアクセス支援

インターネット上での情報資源アクセスを効率的に行うためのポータルサイトあるいはゲートウェイサイトの重要性が広く認められている。実際にこうしたサービスを実現するには、「有用な情報資源」あるいは「高品質な情報資源」を見つけ、それらに関して高品質なメタデータを記述し、適切に組織化し、メンテナンスすることが求められる。こうした機能はなんらかの主題の下に形成されるのでサブジェクトゲートウェイ[7][8]と呼ばれる。たとえば、大学図書館では特定の学術分野や大学等で発信される有用な情報資源に関する情報を収集し、蓄積提供することが行われる。公共図書館では地域の情報資源や児童向きの情報資源に関する情報を収集し、蓄積提供することが多く行われる。

サブジェクトゲートウェイ間の協調によって価値を付加できる。イギリスの Resource Discovery

Network はサブジェクトゲートウェイを結ぶ要の役割を持つもので、ここを介して利用者は有用な情報資源を探ることができる一方、サブジェクトゲートウェイ間の協調によってメタデータ作成のための多重投資など、無駄を省くことができる。

現状ではこれらの仕事の多くの部分を人手に頼らざるを得ず、また、多数の情報資源の組織化を適切に行うための知識も求められる。

(5) ネットワークを介したレファレンスサービス

レファレンスサービスは、図書館員によるサービスであり、図書館の最も特徴的なサービスのひとつである。現在、レファレンスサービスをネットワーク上で提供しようという取り組みが進められている[9]。レファレンスサービスは、質問者が何を聞きたいのかを明確化することなども必要とされる人間主体の作業である。そのため、レファレンスサービスすべてを機械化することは困難であると思える。しかしながら、レファレンスサービスを支えるための環境をネットワーク上に作り上げることや、レファレンスから得られた知識(たとえば、よく質問される項目とその回答、回答に至る過程や有用な資料など)を提供することは重要であろう。

(6) 電子化資料の長期保存

資料の長期保存は図書館の機能としては基本的なものである。ところが、情報媒体技術の進化の激しさは、電子媒体に蓄積された資料、特にもともと電子的に蓄積された資料(Born Digital)の長期にわたる保存を難しくしている[10][11]。アメリカ議会図書館の場合、デジタル資料の長期保存が重要課題として報告されており[12]、また議会からそうした課題を中心としてデジタルライブラリをさらに開発するための大きな予算を与えられプロジェクト NDIIPP が進められている[13]。

(7) 図書館内外におけるデジタルライブラリ機能を利用した総合的な利用環境の整備

現在の大学図書館では、冊子体などの従来型の出版物、電子出版物、ネットワーク上の情報資源を総合的に利用できるハイブリッド環境[14]を作ることが一般的になっている。館内で総合的な利用者環境を構築することに加えて、館外(あるいはキャンパス内外)の利用者に対しても、ネットワークを介した柔軟なサービスを提供して行くことが求められる。

(8) その他

一方、諸外国と比較して我が国を見た場合、図書館とそれ以外の組織、特に研究開発部門との連携が弱いと感じられる。また図書館間での連携も弱いように感じる。研究開発能力を高めるためには必要な情報が容易に手に入る必要があることは言うまで

もない。たとえば、アメリカの場合、Digital Library Federation という組織が早くから作られ、情報の共有やプロジェクトを進める母体にもなっている。我が国を見ると、コンテンツをより使いやすくするための環境整備に十分な手当てがなされていないように感じる。

2.3 デジタルライブラリに関するその他の話題、取り組み

インターネット環境における新しい取り組みとしていくつか興味深いものがあるのでそれを紹介したい。

2.3.1 学術資料のリポジトリの協調

プレプリント、テクニカルレポート、学位論文など電子的に作られた資源を蓄積提供するリポジトリサービスによる協調プロジェクトである Open Archives Initiative (OAI) である[15]。これは別個に作り上げられてきたサービス間でコンテンツを協調的に提供する環境を作り出すことを目指しており、たとえば資源に関するメタデータの収集 (metadata harvesting) のためのプロトコルを決めている。このプロジェクトは、複数のリポジトリによる横断的な利用を可能にするのみならず、収集したメタデータを利用した付加価値サービスの提供を可能にするものである。

上のようなサービスの広がりや学術情報の流通経路を変える可能性を持っている。学術雑誌に掲載される論文が、“品質の保証された資料”として学術情報の中心であることは疑えない。しかしながら、学術雑誌の価格が高く、かつ学術領域の広がりとともに雑誌の数が増えていくことは大学図書館と研究者にとっては頭の痛い問題である。その一方、WWW 上に無料で公開されるテクニカルレポートや学位論文などの学術情報資源はどんどん増えている。査読システムによって内容は保証されるが手に入れにくい(高価な、あるいは限られた場所でしか利用することのできない)雑誌に掲載された論文と、内容は自分で判断しなければならないが WWW 上で容易に手に入れることのできる論文のどちらが本当に役に立つのか、といった問題に関する議論を進めていくことが求められる。

2.3.2 電子化資料の長期保存・アーカイブ

これまで図書館や文書館では紙の資料を中心とした資料保存を行ってきた。先に述べた貴重資料の電子化のように「保存とアクセス」の両面から資料の電子化が進められる一方、電子媒体による出版の発展は電子出版物の保存という困難な問題をもたらしている。これは、納本図書館である国立図書館、我が国の場合は国立国会図書館にとって非常に大きな問題である。

電子出版物は大きく分けて CD-ROM などのパッケージに入れて出版されるものと、ネットワーク上で出版されるものがある。前者をパッケージ系出版物、後者をネットワーク系出版物と呼ぶことにする。パッケージ系、ネットワーク系いずれの場合も、コンテンツの利用に特別のソフトウェアあるいは利用環境が必要とされる場合には、そうしたソフトウェア環境を稼動可能な状態で保存しつづけることが難しい。また、WWW 上で提供される資料の場合、どのような資料を保存の対象とすべきかを定めること(保存対象範囲の定義)が必ずしも容易ではない。電子ジャーナルの場合、印刷物と異なり出版社はライセンスのみを提供することになるので実体の保存に対し誰が責任を持つのかといった問題が生じる。電子資料の保存に関してはコンテンツの内容的問題、知的財産権に関わる問題など、社会制度的な問題もあるが、技術的にも克服しなければならない点が多くある。

電子資料の保存に関して、議会図書館による NDIIPP の取り組み、ヨーロッパの国立図書館による取り組み (NEDLIB)、オーストラリア国立図書館による取り組みなど国立図書館を中心とした取り組みがある。また、議会図書館等とも協調しながら、Internet Archive はインターネット上のコンテンツの収集と保存を進めている。

国立国会図書館では、納本制度審議会が 2004 年 12 月に「ネットワーク系電子出版物の収集に関する制度の在り方について」の答申を提出した[16]。ここでは、従来の納本制度の枠組みとは別の枠組みで収集することとしている。また、ネットワーク系出版物の多様性に合わせた取り組みの必要性について述べている。国による e-Japan 計画の中でも電子資料の保存の問題の重要性は認められており、今後デジタルリソースの保存に関する活動が進むことが期待される。

3. メタデータについて

3.1 メタデータとは

メタデータは簡単には「データに関するデータ (Data about Data)」と定義される。この定義に基づくと、目録、索引、抄録、シソーラスといったものから、資料の識別子、書評や利用条件などまですべてメタデータに含めて考えることができる。利用者はインターネットやデジタルライブラリ上で情報資源を見つけ出し、その内容を評価し、利用する。メタデータなしにはこうした操作を行うことは困難である。

図書館の視点からは、メタデータは従来の図書目録や書誌情報と同じものと(乱暴には)言うこともできる。しかしながら、メタデータは内容とメディアの多様化、利用者や利用環境の変化などに対応するべく提

案されてきているものである。たとえば、図書館での目録は、基本的には、1冊の本、1タイトルの雑誌に対して付与されている。これは、図書館での資料の取り扱いの単位に対応する。一方、インターネット上では、単純にとらえると URI が与えられる資料単位が操作の基本単位ということになり、図書館での資料とインターネット情報資源の違いが良く理解できる。たとえば、議会図書館が開発した MODS (Metadata Object Description Schema) は、MARC21 を基礎にして、多様な目的のために利用できるエレメントセットを定義し、XML での記述形式を定めている[17]。EAD[18]や METS[19]のように電子化資料をアーカイブするために必要な内容情報や構造情報、管理情報を含めて記述するものもある。また、これらは XML による記述形式を決め、データの交換性を高めている。

メタデータに関して、記述の視点と利用の視点の違いに留意する必要がある。記述の視点からは、できるだけ正確かつ詳細に対象資源に関して記述することが望まれる。ところが利用の視点からは、必ずしも詳細な記述は必要ない。一般の利用者による情報資源検索には詳細な情報が利用されることは少ない。異なる規則・基準に基づいて作られたメタデータを横断的に検索する場合や、メタデータをネットワーク上で収集して利用する metadata harvesting の場合には詳細情報は必ずしも用いられない。

現在、応用に応じているいろいろなメタデータ規則が提案されている。たとえば、インターネット上の情報資源の発見を目的として定義され、広い範囲で利用されている Dublin Core Metadata Element Set[20][21]、学習、教育情報資源に関するメタデータである IEEE Learning Object Metadata (IEEE LOM) [22]、政府行政情報[23]、地理情報や環境情報[24][25]、出版物や電子商取引[26][27][28]、資料の保存[29][20][31]などいろいろな分野でのメタデータ開発が進められている。

3.2 メタデータ規則(スキーマ)の利用

3.2.1 メタデータの基本モデル

ネットワーク上には多様な情報資源が提供されると同時に、それらの利用目的、利用方法、利用環境も多様である。このことは、ひとつの情報資源であってもいろいろな視点からメタデータを付与することの必要性を意味する。たとえば、教材として利用するビデオ資料の場合、教育資料としてのメタデータとビデオデータとしてのメタデータの両方が求められる。ネットワーク上での情報資料検索の場合、政府関係資料や地理情報資料というように分野を特化して利用す

る場合はそれぞれの資料向けのメタデータが望ましいが、広い範囲から資料を探す場合は Dublin Core のようなコアメタデータが望ましい。

また、メタデータの記述対象の捉え方も様々である。Dublin Core の場合は、情報資源を対象とし、情報資源を扱う主体は視野の外である。IFLA の FRBR モデル(図 1 参照)では、対象を work, expression, manifestation, item の 4 段階でとらえている[32][33]。OpenURL の概念モデルを与える Bison Futé モデルはコンテキストに基づいて参照先を決定するための基本モデルを与えている[34]。電子商取引用の indecs[35]の場合、たとえば、人、物、取引、権利などが記述対象となる。IMS ではアクセシビリティに関するメタデータエレメントを決めている[36]。RSLP モデルではデジタルコレクションのためのメタデータモデルであり、コレクションの名前や履歴などに加えて、Collector や Owner といったものを記述要素として持っている[37]。

意味的な相互利用性を高めるためにエレメントの定義と記述形式の定義をできるだけ分離して定義してきた Dublin Core を基礎にして、メタデータスキーマのモデルを考えてみたい。Dublin Core のエレメント、限定子、アプリケーションプロファイル(図 2 参照)を整理すると、次の 4 項目に分けることができる。

- (1) 情報資源が持つ属性の名前
- (2) 情報資源が持つ属性の値の記述形式
- (3) 具体的な記述形式にはよらないメタデータ記述の構造定義
- (4) 具体的な記述形式

このようにメタデータ規則を整理することで、記述要素の意味(名前とその役割の定義)と構文的性質を分離して定義できる。複数のメタデータ規則(エレメントセット)から必要なエレメントを選んで応用にあったエレメントセットを作り上げることができる。また、エレメント単位での意味的な関係付けができれば、異なる応用間でのメタデータの相互利用が行いやすくなるといった利点を持っている。

3.3.2 メタデータとメタデータ規則の相互利用

ネットワーク上ではいくつものデータベースに対する横断的な検索が有用である。たとえば、Z39.50 は分散環境上での横断的な検索に広く利用されているプロトコルである。横断検索を実現するには、共通のメタデータ規則の下に検索質問を受け付けることが必要であるため、異なるスキーマで記述される場合、メタデータ規則間あるいはメタデータエレメント間の対応関係の定義が必要とされる。

WWW コンソーシアムによる Resource Description Framework は WWW 上でのメタデータ記述のための共通基盤を提供するための取り組

みである。RDF は Model and Syntax と Schema の二つに分かれている。前者はメタデータ記述のための構文とその基になるモデルを定義し、後者はメタデータの要素や限定子などメタデータ規則に含まれる語彙の定義を与える仕組みを定義している。Semantic Web[38]の取り組みは、WWW 上での情報流通のための意味的基盤を与えようというものである。現時点での中心は、メタデータのための語彙を含め、WWW 上での情報交換のための語彙の定義 (ontology あるいは vocabulary) が中心的話題となっているように思える。Dublin Core においても、メタデータを記述するための統制された語彙に関する議論は多くあり、こうした問題は古くて新しい問題の一つのようである。

ネットワーク上でいろいろなメタデータ規則が用いられる。Dublin Core の範囲内だけで見てもいろいろな拡張がなされている。メタデータの流通性、相互利用性を高めるには、メタデータ規則に関する情報

の流通性を高めることが求められる。そのため、メタデータ規則を登録し、提供するサービスであるメタデータスキーマレジストリ(あるいは、メタデータレジストリ)の重要性が認められている。DCMI でもレジストリに関するワーキンググループを作り、DCMI レジストリの構築を進めており、筆者等も DCMI と協調してレジストリの構築を進めている[39][40]。メタデータレジストリは、単に要素や限定子の定義の参照記述を人間が見るためだけに限らず、要素や限定子に対し、ネットワーク上で一意に定まる識別子を与えることで要素などの意味定義の基底を与えること、要素などを長期にわたって維持管理するための基盤となることが期待される。さらに、新しい要素や限定子を定義する上で参考となる情報を与えるといった付加価値サービスを行うことができると考えられる。

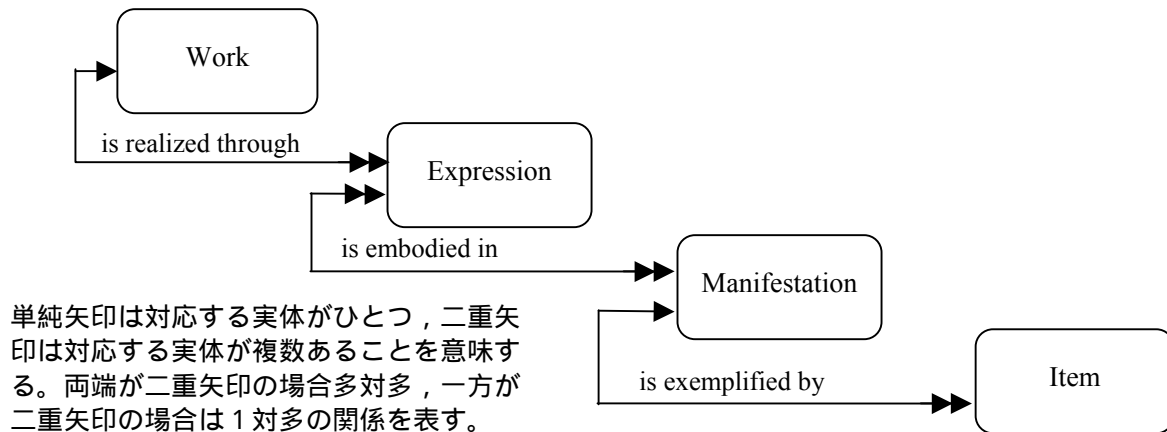
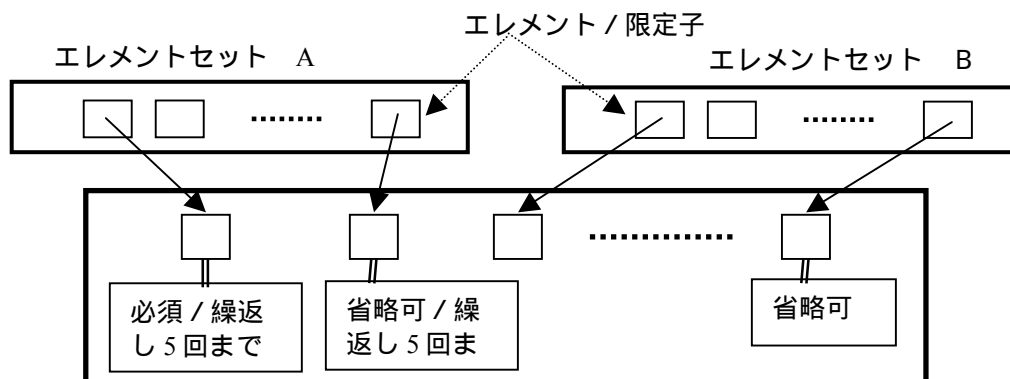


図1 IFLA FRBR における知的内容を表す実体 (Group 1 Entity) 間の関係



ある応用のために作られたアプリケーションプロファイル

図2 エlementセットとアプリケーションプロファイル

4. デジタルアーカイブとメタデータについて

デジタルコンテンツのアーカイブを実現するには、保存すべき資料を収集し、コンテンツの利用性を保ちながら蓄積・組織化し、利用者に提供することが求められる。

4.1 デジタルコンテンツの収集

図書館や文書館では、館毎に決められた資料の収集方針に基づき、収集対象範囲を決め、資料を収集する。そして、資料の保存方針に従って、収集資料の保存期間や保存形態を決める。したがって、コンテンツの収集はアーカイブの最初の段階である。パッケージ系コンテンツの場合は、コンテンツを納めた「もの」を収集すればよいが、ネットワーク系コンテンツの場合には、自動収集の技術、収集範囲と対象の指定、収集対象の内容評価といった問題がある。

(1) 収集方針と自動収集の技術

収集方針を大別すると、対象範囲のコンテンツを網羅的に収集しようというものと対象範囲の中から保存すべきものを選択的に収集しようというものがある。網羅的収集の場合には、ソフトウェアによる自動収集を行うのが一般的である。選択的収集の場合には、人的に収集対象コンテンツを選定し収集することが行われる。このほか、出版者から保存すべきコンテンツを収集者に提供する場合や出版者と収集者の合意に基づき限定的に自動収集する場合もある。選択的に収集する場合は、コストが高かつき、またカバーできる範囲が限られるといった問題がある。一方、自動収集する場合、収集ソフトウェアのアクセス可能範囲の問題に加えて、ハイパーテキスト構造や動的構造を持つ電子文書に起因する問題がある。

(2) 収集範囲と対象の指定

自動収集の場合、ドメイン名や IP アドレスといったアドレス情報、テキストやイメージといったコンテンツタイプの識別情報を用いて収集対象範囲を限定することが行われている。しかしながら、保存を目的とした収集の場合、コンテンツの内容によって収集範囲と対象を指定することが求められる。

(3) 収集対象の内容評価

自動収集の場合、コンテンツの内容が収集方針に合致したものであるかの判断を自動化することが求められる。また、収集したコンテンツを公開するには、非合法コンテンツやプライバシーに関わるコンテンツなど公開に適しないものを公開対象から除外する必要が生じる。また、コンテンツの内容に関する信頼性のように判定の自動化の困難な問題もある。

4.2 デジタルコンテンツの保存

コンテンツの保存は、基本的には、「もとのままの状態を利用できること」を意味する。しかしながら、図書

館では印刷物をマイクロフィルム化して保存するなど、もとのままの状態にこだわらず、保存方針に合わせてコンテンツの内容を利用可能な状態で保存することも行われてきている。デジタルコンテンツに関しても同様のことが言える。たとえば、特定のブラウザに合わせてデザインされたウェブページの場合、保存対象となっているページのソーステキストだけを保存するか、あるいはページの Look and Feel までも保存するためにそれを閲覧するためのブラウザおよびブラウザを動作させるためのハードウェアとソフトウェア(これをコンテンツの利用環境もしくは単に環境と呼ぶ)もあわせて保存するかといった点がある。こうした点は、保存方針によって決められるべきことである。また、「長期保存」が何年くらいを意味するのかについては明確な定義はない。コンテンツの種類や保存の目的によっても異なる。

保存の単位は、物理的単位と論理的単位の両方でとらえる必要がある。物理的単位としてはファイル単位あるいはデータベース単位がある。論理的単位には、1 点の電子文書あるいはデータ、資料の集まりとして構成されるひとまとまりのコレクションがある。前者の例にはワープロで作成された電子文書、Web 文書、物理資料を電子化することで作られた資料などがある。後者の例には、いくつもの論文や記事を含む 1 タイトルの雑誌、いくつもの WWW ページを集めて作るコレクションなどがある。コレクションの場合は、コレクションに含まれる資料をも含めて保存すると考えるのが自然である。

4.3 デジタルコンテンツの保存のためのメタデータ

4.3.1 OAIS 参照モデルに基づくメタデータスキーマ

OAIS 参照モデルでは、情報オブジェクト (Information Object) がデータオブジェクト (Data Object) とその表現情報 (Representation Information) からなるとしている。データオブジェクトは、具体物 (物理オブジェクト, Physical Object) あるいはビットの集まりによって生成されるデジタルオブジェクト (Digital Object) である。表現情報も何らかのデジタル表現された実体である。OAIS 参照モデルでは、情報オブジェクトを保存するために情報パッケージの概念を導入している。情報パッケージは図 3 に示すように情報オブジェクトを内容情報 (Content Information) として保存するため、保存記述情報 (Preservation Description Information, PDI) とあわせてパッケージ化している。また、パッケージ化したオブジェクトの保存に際しては、情報パッケージに関する記述 (Descriptive Information about Package) とする。PDI は保存対象コンテンツに関する情報を持つものに対し、他はパッケージに関する情報をもつ。PDI の内容は次の 4 種類の情報に規定されている。

- ・ 来歴 (Provenance) : コンテンツの出処やコンテンツの保存に関わる処理の履歴。
- ・ コンテキスト (Context) : 情報パッケージ外のオブジェクトとの関係。たとえば、コンテンツが作られた理由、他のコンテンツとの関係。
- ・ 参照 (Reference) : コンテンツを一意に識別するための識別子、もしくは識別のためのシステム。たとえば、ISBN やコンテンツを一意に識別するための属性の集合。

- ・ 不変性 (Fixity) : コンテンツを保護するための情報。たとえば、チェックサム。
デジタルコンテンツの保存のためのメタデータスキーマに関する検討が進められ、Cedars や NEDLIB、OCLC (Online Computer Library Center) と RLG (Research Library Group) の共同ワーキンググループでは OAIS 参照モデルに基づくメタデータスキーマスキーマの報告を出している[29][30][31]。

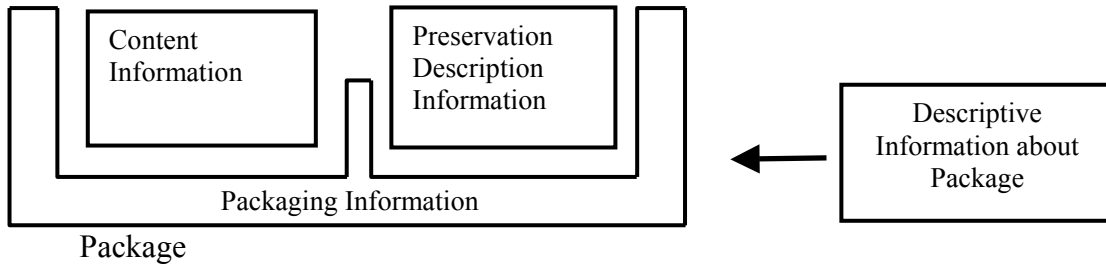


図 3 OAIS の Information Package の概念構造

4.3.2 Metadata Exchange and Transmission Standard (METS)

METS は Encoded Archival Description (EAD) などでの経験をもとに、Digital Library Federation を中心に作られたデジタルコンテンツのためのメタデータ記述の標準である。METS は OAIS 参照モデルにおける情報パッケージ (SIP, AIP, あるいは DIP) として用いることができる。METS の記述には下に示す要素が含まれており、また XML によるメタデータ記述の形式を決めている。

- ・ Header: XML テキストとして内容の属性を表すヘッダ。
- ・ Descriptive Metadata: コンテンツの検索等のために利用する記述的メタデータ。外部の記述的メタデータ(たとえば、MARC レコードなど)を参照する場合もある。
- ・ Administrative Metadata: ファイルの生成と管理の記録、知的財産権情報、デジタルオブジェクトの作成の対象となった、もとの非デジタルオブジェクト、デジタルオブジェクトの来歴情報など管理のための記述。
- ・ File Groups: デジタルオブジェクトを構成するひとつもしくは複数のファイルをグループ化したもの。同一の内容を異なる方式によって作成したファイルのリスト。
- ・ Structural Map: デジタルオブジェクトの構造の記述。たとえば、セクションに分かれた録音データの場合、全体とセクションの関係。

- ・ Structural Links: Web ページのようなハイパーテキスト構造を持つ資源のリンク情報の記述。
- ・ Behavior: オブジェクトが持つ動的性質の記述。

6. おわりに

国内の図書館現場でのデジタルライブラリサービスはかなり進んだと思われる。たとえば、大学図書館で提供される電子ジャーナルのタイトル数も最近では非常に増加した。図書館(あるいはそれに似た機能を持つ組織)はネットワーク上での出版、情報流通の増加に対応した中継ぎとしてのサービスを提供することが基本であるという視点からはずいぶん進歩したと感じる。

コンテンツの提供のほか、図書館に求められる機能として重要なものに、サブジェクトゲートウェイやネットワーク上でのレファレンスサービスなど情報資源へのアクセス支援サービスがある。こうしたサービスを支える技術が十分には開発されていないので、人手中心に開発が進められている。資料の長期保存技術とともにこうした情報技術の開発が望まれる。また、近年の携帯電話の急速な発展と普及により情報アクセスのための利用者環境の多様化が進んでいる。一方、BLOG などの情報発信や収集の環境、Web サービスといったサービス提供のためのインタフェースも進化している。デジタルライブラリのサービスをこうした環境変化に合わせていくことも求められる。

デジタルライブラリにおいて重要と思われる点を、メタデータの視点を中心に、いくつか挙げてみたい。

(1) 多様な interoperability とメタデータ規則の流通： ネットワーク上では、それぞれの応用にあわせて規則を決めることのできる自由度が求められる一方、地理的距離、時間的距離、言語や文化の違い、分野の違いを超えてネットワーク上で相互にデータやサービスを利用し合うための共通性が同時に求められる。そこではいろいろなレベルでの相互利用性が求められることになる。同時に、そうした interoperability を支えるにはメタデータ規則(メタデータスキーマ)をネットワーク上で流通させる仕組みが重要である。

(2) メタデータの効率的な作成と維持： 現時点では高品質であることを求められるメタデータの作成には人手がかかる。また、そうしたメタデータを作成する対象を選定すること、すなわちある程度のコストをかけてもメタデータを作る価値がある情報資源であると判断することにもコストがかかる。加えて、ネットワーク情報資源の特性として内容の更新が容易であるという点がある。この点はメタデータの維持管理を難しくしている。こうした点に対する情報技術開発が望まれる。

(3) 多様な利用者のための利用性： 障害者を含めた多様な利用者のための柔軟な利用環境を提供すること、多言語情報アクセスを実現すること、利用環境に応じたコンテンツ提供を行うことなど、多様な利用者、資源、利用環境に対応するためのメタデータが求められる。

デジタルライブラリにとって、作成したデータ(メタデータを含む)をネットワーク上で相互に利用できるようにすれば、「館」に縛られたサービスとは大きく異なるサービスを提供できる。コンテンツサービスの集中化と利用者サービスポイントの分散化を進めることができる。コンテンツサービスの集中化が可能であるとは言っても多様な利用者のニーズに 1 つのサービスで応えることは困難であるので、いくつかのサービス間の協調が求められることになる。この視点からはデジタルライブラリサービス間の相互利用性は非常に重要な問題である。コンピュータ上に実現されるシステムのレベルでの相互利用性を高めることが重要であることは疑えないが、その一方、個々のサービスの自由度を高めることとの両立は必ずしも容易ではない。そのため、人間をも含めたシステムとしての相互利用性を高めるためのサービス間の協調が求められる。

図書館サービスは情報の発信者(出版社など)と受信者(図書館利用者)の間に位置付けられる付加価値サービスである。ネットワークによって発信者と受信者が直接結び付けられるとは言え、中間的なサービス、たとえば、提供される情報の品質の評価や情報への高い検索性やアクセス性の保証、知的財産権の保護やプライバシーの保護などの付加価値サービス無しにはネットワーク上に提供される様々なコンテンツを十分に使いこなすことはできない。また、こうした付加価値サービスの実現には、ひとつの図書館単独での努力では限界があり、図書館間および図書館と関連組織間での協調が求められる。

最後に、デジタルライブラリには知的財産権やプライバシーなど社会制度とも関連する問題が多くある。これらは筆者の能力の範囲外であるのでここでは触れていない。これらはケース・バイ・ケースでの解決が求められることも多く、デジタルライブラリとしての経験を重ねることとそれを共有することが求められるようである。

参考文献(下記の文献のうち URL の記載したものは、2005 年 5 月時点で存在を確認済み。)

- [1] Digital Library Initiatives, <http://www.dli2.nsf.gov/>
- [2] National Science, Mathematics, Engineering and Technology Education Digital Library, <http://www.nsdlib.org/>
- [3] American Memory: Historical Collection for the National Digital Library, <http://memory.loc.gov/>
- [4] D-Lib Magazine, <http://www.dlib.org/>
- [5] Ariadne <http://www.ariadne.ac.uk/>
- [6] 歳森, 宇陀, 松林, 電子情報環境下における科学技術情報の蓄積・流通の在り方に関する調査研究 2004.7, 107p, http://www.ndl.go.jp/jp/library/lis_research/lis_rr_02.pdf
- [7] 緑川他, サブジェクトゲートウェイ: ネットワーク上の知識集積, 知的コミュニティ基盤研究センター モノグラフシリーズ 1, <http://www.kc.tsukuba.ac.jp/monograph/monograph01.pdf>
- [8] 緑川他, 「サブジェクトゲートウェイ 2: 知識集積の組織化」 知的コミュニティ基盤研究センター モノグラフシリーズ 3, <http://www.kc.tsukuba.ac.jp/monograph/monograph03.pdf>
- [9] Global Reference Network, <http://www.loc.gov/rr/digiref/history.htm>
- [10] Hedstrom, M. and Ross, S., "Invest to Save, Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation", [- 67 -](http://delos-</div><div data-bbox=)

- noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf, 2003
- [11] 栗山正光, “長期保存型電子図書館とOAIS参照モデル”, 公開シンポジウム「電子図書館の軌跡と未来」論文集, 筑波大学附属図書館, 2003.1
- [12] Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, National Research Council, LC21: A Digital Strategy for the Library of Congress, 238p., <http://books.nap.edu/html/lc21/>, 2000
- [13] National Digital Information Infrastructure and Preservation Program, <http://www.digitalpreservation.gov/>
- [14] 永田治樹, “ハイブリッド図書館のビジネスアーキテクチャ, 公開シンポジウム「電子図書館の軌跡と未来」論文集, 筑波大学附属図書館, 2003.1
- [15] Open Archives Initiative, <http://www.openarchives.org/>
- [16] 答申「ネットワーク系電子出版物の収集に関する制度の在り方について」, 2004.12, http://www.ndl.go.jp/jp/aboutus/data/a_toushin_2.pdf
- [17] MODS: Metadata Object Description Schema (official web site), <http://www.loc.gov/standards/mods/>
- [18] Encoded Archival Description (EAD) – Official EAD Version 2002 Web Site, <http://lcweb.loc.gov/ead/>
- [19] METS: Metadata Encoding & Transmission Standard (official web site), <http://www.loc.gov/standards/mets/>
- [20] Dublin Core Metadata initiative, <http://dublincore.org/>
- [21] 杉本重雄, “Dublin Core について (2 回連載)”, 情報管理, Vol.45, no.4, 2002.7, pp.241-254, no.5, 2002.8, pp.321-335
- [22] Learning Technology Standard Committee, <http://ltsc.ieee.org/>
- [23] GILS: Global Information Locator Service, <http://www.gils.net/>, http://www.access.gpo.gov/su_docs/gils/index.html
- [24] metadata, <http://www.fgdc.gov/metadata/metadata.html> (Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998))
- [25] Alexandria Digital Library Project, <http://www.alexandria.ucsb.edu/> (ADEPT Metadata Standard, <http://piru.alexandria.ucsb.edu/metadata/>)
- [26] Digital Object Identifier, <http://www.doi.org/>
- [27] ONIX Product Information Standard, <http://www.editeur.org/onix.html>
- [28] Indecs Home Page (Interoperability of Data in a e-Commerce Systems), <http://www.indecs.org/>
- [29] Cedars Project, “Cedars Guide to Preservation Metadata”, 2002, <http://www.leeds.ac.uk/cedars/guideto/metadata/>
- [30] Lupovici, C., Masanes, J., “Metadata for Long Term Preservation”, NEDLIB Report series, no.2, 2000, <http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf>
- [31] The OCLC/RLG Working Group on Preservation Metadata, “Preservation Metadata and the OAIS Information Model – A Metadata Framework to Support the Preservation of Digital Objects”, 2002, http://www.oclc.org/research/pmwg/pm_framework.pdf
- [32] IFLA, “Functional Requirements for Bibliographic Records”, <http://www.ifla.org/VII/s13/frbr/frbr.pdf>, 1998
- [33] 和中, 古川, 永田, 書誌レコードの機能要件—IFLA 書誌レコード機能要件研究グループ最終報告, 121p, 2004
- [34] Van de Sompel, H., and Beit-Arie, O., “Generalizing the OpenURL Framework beyond References to Scholarly Works: The Bison-Futé Model”, D-Lib Magazine, vol.7, no.7/8, 2001, <http://www.dlib.org/dlib/july01/vandesompel/07vandesompel.html>,
- [35] Rust, G., “The <indecs> metadata framework”, <http://www.indecs.org/pdf/framework.pdf>, 2000.6
- [36] IMS Global Learning Consortium: Accessibility Specification, <http://www.imsglobal.org/accessibility/>
- [37] Powell, A. et al., “RSLP Collection Description”, D-Lib Magazine, vol.9, no.9, 2000, <http://www.dlib.org/dlib/september00/powell/09powell.html>
- [38] Semantic Web, <http://www.w3.org/2001/sw/>
- [39] DCMI Registry Working Group, <http://www.dublincore.org/groups/registry/>
- [40] Nagamori, M., Sugimoto, S., A Metadata Schema Framework for Functional Extension of the Metadata Schema Registry, Proceedings of DC-2004, pp.3-11, 2004.10, Shanghai, China