

1. はじめに¹

Digital Library²(ここではデジタルライブラリあるいは DL 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地的な DL 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]。DL 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、図書館で直接そのまま利用できるシステムの開発を目指したものとは言えないが、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、NSF による National SMETE Digital Library(NSDL)³プログラムは、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ている[2]。これは教育学習コンテンツの開発をもサポートするものであり、インターネット上に教育学習コンテンツとそ

の利用環境(すなわち図書館)の実現を目的とするものである。

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。図書館による取り組みとして典型的なものには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がある。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。

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作り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。それらのひとつに電子資料、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があり、また Interoperability がある。

図書館にとって資料の保存が重要であることは言うまでも無いが、新しい情報技術を使って作られる資料は、技術の変化とともに利用ができなくなるため、保存の観点からは大きな問題を抱えている。国立図

¹ 本稿は筑波大学附属図書館主催電子図書館シンポジウム(2003年1月24日、<http://www.tulips.tsukuba.ac.jp/pub/dlsympo/>)でのチュートリアル資料[59]および拙著工知能学会誌記事[60]に加筆したものである。

² 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。また、Electronic Library、仮想図書館や壁なし図書館などのことばも同一の意味と捉える。一方、「図書館」ということばから、従来の建物ないしは施設を基盤としたサービスを連想させることが多い。こうした「館」ということばからの連想を避け、機能を中心に述べるには「館」を避けた記法が良いと思われるため本稿ではデジタルライブラリあるいは DL と記している。

³ SMETE は Science, Mathematics, Engineering and Technology education に対応する。

書館を中心として電子資料の保存に関するいろいろな取り組みが進められている。

Interoperability がインターネット上の情報流通、DL にとって重要であることは言うまでもない。たとえば、1997-98 年に DL の研究戦略を議論するワーキンググループでも重要な研究トピックとして議論されている[4]⁴。また、2001 年の Joint Conference on Digital Libraries (JC DL 2001) での基調講演において、Clifford Lynch 氏は DL における interoperability について述べた⁵。そのなかでは Z39.50 のようなプロトコルによる図書館システム間の interoperability だけではなく、Dublin Core における Dumb-down 原理のような「ゆるい interoperability」についても述べられていた。さまざまな草の根的活動から多様な情報資源が作り出されるインターネット環境においては、標準規格に基づくトップダウンで厳密な相互利用性だけではなく、ボトムアップでより緩やかな相互利用性が求められると考えられる。また、WWW コンソーシアムによる Semantic Web の活動は、文字コードやテキストのレベルから XML によって構造化テキストのレベルまで上がった interoperability のレベルを、さらにより意味的なレベルに引き上げようというものであると理解できる。

本稿では、これまでの取り組みを振り返りながら、現在のデジタルライブラリに関する概説を行うとともに、今後の課題について述べる。以下、第 2 節で 1990 年代に進められた DL に関するいろいろな研究開発活動について概観し、簡単にまとめる。第 3 節では、Dublin Core の解説とメタデータに関して今後求められると考えられる視点について述べる。第 4 節では電子資料の保存に関して紹介する。第 5 節では interoperability に関して簡単に述べる。最後に、これらをまとめて考察する。

⁴ DLI2 の期間にも NSF-DELOS で幾つかのトピックに関する議論を行うワーキンググループが作られている。トピックとしては Spoken-Word Digital Audio Collections, Information Extraction from Digital Libraries, Personalization and Recommender Systems in Digital Libraries, ePhilology: Emerging Language Technologies and Rediscovery of the Past, Digital Imagery for Significant Cultural and Historical Materials, Preservation and Archiving がある。

参考: <http://delos.noie.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/joint-wgs.html>

⁵ アメリカ, Coalition for Networked Information of the Executive Director. 講演題目: Interoperability: the Still-Unfulfilled Promise of Networked Information

2. Digital Library

2.1 DL のための新しい技術の研究開発の活動

1990 年代にはいくつかの DL 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 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⁶。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。

NSDL は高等教育におけるデジタルライブラリのサービス基盤からコンテンツまでをカバーする研究助成プログラムである。一方、インターネットの教育利用に関しては、K-12⁷として広く進められてきており、大学学部レベルでの教育とあわせて K-16 といわれることもある。現在、あちこちの大学で遠隔教育・学習プログラムの提供が進んでおり、インターネットを利用した学習、教育は重要なトピックとなっている。

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。DLI-1 では、NSF のほか、DARPA と NASA が研究助成母体として参加した。一方、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 DL 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。また、デジタル資料の長期保存といった図書館の基本機能に関するものは DLI-2 で開始された。

2.2 図書館を基礎としたデジタルライブラリの開発

⁶ これからはじまる第 6 フレームワークでも IST での研究助成は続けられるようである。

⁷ K-12 は Kindergarten to 12th Grade を意味し、幼稚園から高校までの教育におけるインターネット利用が進められている。

従来、研究は大学や研究所に所属しないと困難であるとされてきた。研究に必要な資料を持つ図書館が近くにないと研究が進められない、というのがその理由の一つである。ネットワーク上に提供される学術情報が飛躍的に増えたとはいえ、図書館(あるいは図書館のようなサービス)が研究者にとって重要であることは、従来とまったく変化していない。しかしながら、インターネット上での情報流通の発達によって、図書館に行く前に(あるいは行かずに)、インターネットで情報を探すということが一般化してきている。行動パターンの違いは分野や個人に依存するところが大きいことは言うまでもないが、こうした傾向は先端的な自然科学・技術分野でより顕著であるように感じられる。また、ネットワークの発展と利用の広がりによるこうした影響は大学図書館だけではなく、公共図書館についても同様なことが言える。

こうした環境下、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 DL 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(典型的なものは電子雑誌やデータベース)
- (2) 貴重資料や歴史資料の電子化と提供
- (3) 大学など図書館が所属する組織・機関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
- (4) インターネット上の情報資源へのアクセス支援
- (5) ネットワークを介したレファレンスサービス
- (6) 電子化資料の長期保存
- (7) 図書館内外における DL 機能を利用した総合的な利用環境の整備

以上の項目に関して簡単に考察したい。

(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。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。小規模の学会などでは、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の電子化と提供

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。この機能は電子博物館や電子美術館とよばれるものと同じと考えても良い。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観点からは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はなく、教育利用での利用性の促進や評価、DLI-2 に見られるような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。情報アクセスの支援は、大学などの組織が発信する情報への入り口(ゲートウェイ、ポータル)を作るものであるといえる。

(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と呼ばれる。たとえば、大学図書館では特定の学術分野毎に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。

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[5][6]。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[7][8][9]。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[10]、また議会からそうした課題を中心としてDL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が進められている[11]。

(7) 図書館内外におけるDL機能を利用した総合的な利用環境の整備

図書館内の計算機とネットワークの環境は、この10年で大きく変わってしまったように思われる。大学図書館では、端末用のパソコンがずらっと並んでいることはもう珍しくない。現代の図書館には、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な環境を作ることが求められる[12]。また、図書館内外(あるいはキャンパス内外)での利用環境の違いを少なくしていくための技術や制度が求められると思われる。

(8) その他

全体を通して、この10年間でずいぶん図書館および利用者の情報環境は変化したことは間違いない。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な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サービス(リポジトリ)による協調プロジェクトであるOpen Archives Initiative(OAI)である[13]。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でき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。研究者にとっては学術雑誌に論文を掲載することは自身の業績評価につながる。その一方、研究者の大学の図書館がその雑誌を購入していない、あるいは購入の継続を打ち切ってしまう自分の論文が掲載された雑誌を図書館で読めない状況も出てくる可能性がある⁸。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。したがって、査読システムによって内容は保証される

⁸ 図書館における電子出版物(特に電子ジャーナル)の導入、提供などに関する種々の問題に関して知識と情報を共有し、意見交換をする取り組みが進められている。たとえば、アメリカを中心とする国際的な取り組みであるSPARC(the Scholarly Publishing and Academic Resource Coalition, <http://www.arl.org/sparc/>)、我が国での国立大学図書館協議会による取り組みなどがある。図書館での現実的な対応に関してはここでは触れない。

が手に入れにくい(高価な、あるいは限られた場所でのみ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観点から、これまで以上に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な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。次に、電子資料の保存に関するいくつかの活動を紹介する。なお、電子資料のアーカイブについては 4 節で詳しく述べる。

- (1) ミシガン大学とリーズ大学(イギリス)の共同プロジェクト CAMiLEON は Emulation は電子資料の保存を主目的としたプロジェクトである[14]。また、コーネル大学のプロジェクト PRISM でも電子化資料の保存を扱っている[15]。
- (2) スタンフォード大学を中心として進められているプロジェクト LOCKSS (Lots of Copies Keep Stuff Safe) は、電子ジャーナルの保存のためにジャーナルへのアクセスのために用いられる PC 上

にコピーを残す仕組みを作り、それをネットワーク上の多数の PC で動作させることによって、実質的に資料を保存するという取り組みを進めている[16]。

- (3) 北欧、オーストラリア他の国立図書館において、WWW 上の出版物を収集し、保存する取り組みを進めている。図書館によって、内容による選択的収集を行っている館と、内容によらない網羅的収集を行っている館がある。

- (4) Internet Archive は、米国議会図書館の補助を受け、インターネット上の資料のアーカイブ作りを目的として WWW 上の資料をロボットで収集し、蓄積・提供するサービスを進めている[17]。ここでは、たとえば「1999 年の首相官邸のページ」を探すことのできるサービス Wayback Machine を提供している。我が国の国立国会図書館では WWW 上の資料を収集し、蓄積提供の実験事業 WARP を開始している[18]。こうしたサービスはそれ自体が WWW 情報資源の保存という観点から重要であるのみならず、WWW 上では削除されてリンク切れとなったページに対してもアクセスを保障してくれるので、サブジェクトゲートウェイのように情報資源に関する情報を提供するサービスにとっては有用なサービスである。

2.3.3 その他

JSTOR[19]は主要な雑誌を遡及的に電子化し、提供するサービスである。雑誌を網羅的に収集することは大きな図書館であっても完璧なものにすることは難しい。JSTOR は雑誌タイトル毎に創刊号から雑誌記事を網羅的に蓄積提供するものであり、個々の図書館だけでは実現しにくい網羅性を実現している。

先に述べたサブジェクトゲートウェイは、学術分野や利用者対象の関心分野に特化して構成される。一方、サブジェクトゲートウェイ間の協調を図ることで広い分野をカバーすることができる。イギリスで進められている Resource Discovery Network[20]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すことができる一方、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。EU の助成の下で進められたプロジェクト Renardus[21]もサブジェクトゲートウェイ間の協調を進めた。

ネットワーク上の有用な情報資源に関するメタデータ作成の取り組みとして、OCLC の CORC や情報学研究所による学術コンテンツポータル[22]がある。はメタデータ作成を分担・協調して行う取り組みである。これらは有用な情報資源の見つけ、そのメタデータを各作業の参加者(参加館)間の協力によって行うものである。

⁹ 国立国会図書館では納本制度に基づくパッケージ系電子出版物の収集を始めているが、ネットワーク系出版物に関しては範囲を限った収集をおこなっているのみである。ネットワーク系出版物に関しては納本制度審議会に小委員会を設け、議論を進めている。

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、書評や内容の Rating 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 DL 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、利用者と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[23]。EAD[24]や METS[25]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) 場合には詳細情報は必ずしも用いられない。

現在、応用に応じていろいろなメタデータ規則が提案されている。その一例を下に示す。

- Dublin Core Metadata Element Set: 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たもの[26][27]。

- IEEE Learning Object Metadata (IEEE LOM): 学習、教育情報資源に関するメタデータ。

IMS, Ariadne など遠隔教育・学習分野で利用されている[28][29][30][31]。

- MPEG-7: オーディオ・ビデオデータのためのメタデータ[32][33]

他に、政府行政情報[34][35]、地理情報や環境情報[36][37]、出版物や電子商取引[38][39][40]、倫理的内容[39]、資料の保存[42][43][44]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¹⁰。

3.2 Dublin Core Metadata Element Set

Dublin Core Metadata Element Set (DCMES) はインターネット上での情報資源の発見を目的として開発が進められてきたもので、現在いろいろな分野で応用が進められている。図書館や美術館・博物館での利用、教育情報資源への適用などが進められてきているほか、ヨーロッパ各国での政府情報のためのメタデータとして利用されている¹¹。

Simple Dublin Core については、欧米を中心として標準規格化が進められ ISO での標準化が行われた¹²。我が国でも大学図書館を中心として利用が進んでおり、国立国会図書館、国立情報学研究所でもネットワーク情報資源のためのメタデータ規則として適用している[22][45]。ここでは、Dublin Core の基本的な考え方と利用の現状について述べる。詳しくは、拙著記事[27]を参照していただきたい。

3.2.1 基本概念

(1) 「情報資源の発見」のためのメタデータ

DCMES は情報資源を発見するためのメタデータ、すなわち利用者が情報資源を探すための情報資源の記述データを提供することが本来の目的である。「情報資源の記述」という観点から見ると、できるだけ詳細に定義されたメタデータエレメントや厳密な意味定義のなされた統制語彙に基づく記述が望ましいと思われる。しかしながら、利用者の観点からは、情報資源を探すために詳細な記述を用いることは多くない。そのため、情報資源の発見の観点から、できるだけ基本的な内容のエレメントからなるメタデータ記述規則を決める努力がなされてきた。

¹⁰ メタデータに関して IFLA が豊富なリンクを提供している[61]。

¹¹ DCMI のホームページからプロジェクトや政府情報のためのメタデータとしての採用例を見ることができる。DC-2001 では環境情報、食料情報などの分野での Dublin Core の利用が報告されている。

¹² ISO では 2003 年 4 月に ISO 15836 として承認した。(http://www.niso.org/international/SC4/n515.pdf, Information and documentation — The Dublin Core metadata element set, 2003-02-26)

(2) コア・メタデータ・エレメント・セット (Core Metadata Element Set)

筆者はコア (Core) ということばが大きな意味を持っていると理解している。ネットワーク上には様々な分野の、様々な種類の、様々な目的で作られた情報資源がある。これらに共通に適用できるメタデータ記述規則を作るには、それぞれの分野や目的で必要とされるエレメントを全て含むように定義する方法と、全ての分野や目的で共通に利用可能であると考えられるエレメントのみを含むように定義する方法がある。前者の方法をとると巨大な規則を作ることになり、実現性に疑問がある。Dublin Core は後者の方法をとったものである。

(3) Simple Dublin Core (Simple DC) と Qualified Dublin Core (Qualified DC)

Simple Dublin Core は、表1に示す15エレメントの定義に加えて、すべてのエレメントが省略可能かつ繰り返し可能という定義を持つメタデータ規則である。一方、この開発過程で、詳細な記述のための限定子 (qualifier) に関する議論が行われてきた。この過程において、「エレメントをより詳細に表す」ことに関する基本的な性質の違いが明確化されて、それに基づき限定子の導入が行われてきた。

表 1 Simple Dublin Core の 15 エレメント

タイトル (Title)
内容記述 (Description)
日付 (Date)
資源識別子 (Identifier)
関係 (Relation)
作成者 (Creator)
公開者 (Publisher)
資源タイプ (Type)
情報源・出处 (Source)
対象範囲 (空間的・時間的) (Coverage)
主題およびキーワード (Subject)
寄与者 (Contributor)
形式 (Format)
言語 (Language)
権利管理 (Rights)

たとえば、作成者を詳細 (あるいは正確に) 表す方法として、正確な記述のための典拠ファイルのレコード名など作成者に与えられたユニークな識別子で記述する方法と、作成者の名前、所属、連絡先などの組で記述する方法がある。前者の場合にはどのような典拠ファイルを使っているのか、あるいはどのような統制語彙を用いて表しているのかを示す必要がある。後者の場合、作成者である人 (あるいは組織) を

表す値の構造、すなわち下位のエレメントの定義が必要である。このほか、「詳細に表す」ということには、「作成者」の意味が小説や論文の「著者」であるのか、「作曲家」あるいは「作詞家」であるのかといった、エレメントの意味の詳細化の場合がある。Dublin Core では値の表現形式に関する指定 (すなわち統制語彙やデータ形式) を指示するもの (Encoding Scheme qualifier, コード化スキーム限定子) と、エレメントの意味の詳細化を表すもの (Element Refinement qualifier, エレメント詳細化限定子) の2種類の限定子を定義している。一方、構造の表現に対応するものは現在の DCMES の限定子としては含まれていない。この理由については後述する。このようにして定義された限定子を含む DCMES が Qualified DC と呼ばれる。

(4) Warwick Framework

1996 年春にイギリスの Warwick (ウォーリック) で開催された第 2 回のワークショップで提案された Warwick Framework は Dublin Core の位置付けとネットワーク上でのメタデータのとらえかたの上で大きな役割を果たしている。Warwick Framework は複数のメタデータ規則による記述 (パッケージと呼ぶ) をひとつの入れ物 (コンテナ) に入れ、まとめるというものである。MARC 等の従来の目録規則では、それ自身で閉じたものとして定義されてきたが、インターネット上での情報資源や利用方法の多様性に対応するには、単一の規則のみで記述することをせず、いくつかの記述規則を組み合わせでひとつのメタデータを記述する枠組みを与えている。この考え方は、WWW コンソーシアムの Resource Description Framework (RDF) [46] の考え方に影響を与えている。

3.2.2 Dublin Core の利用

(1) Dublin Core の適用方法

現在、DCMI 推奨エレメントは Simple DC の 15 エレメントに Audience エレメントを加えた 16 エレメントである。これに DCMI 推奨限定子、および DCMI Type Vocabulary を加えたものが推奨記述要素セットである。一方、これまで Dublin Core はいろいろな応用に利用されてきている。こうした応用では、推奨エレメント・限定子に加え、それぞれの目的に則して独自の拡張がなされていることが多い。また、Open Archives Initiative に見られるように、異なるメタデータ規則の下に作られたメタデータの共通部分として Dublin Core を利用することもある。一般に、メタデータの記述の際にはできるだけ「高品質な記述」が求められる。「高品質な記述」には、できるだけ詳細であること、できるだけ内容の記述が統制されて

いることが含意されている。一方、前にも述べたように情報資源の発見の側からは、あまりに厳密な規則はエンドユーザには意味のないものである。

(2) Dublin Core の具体的な記述形式

DCMI では具体的な記述形式の定義を HTML や XML 以外には積極的には進めてこなかった。その理由は、記述形式の定義を進めることによってエレメントの意味定義に影響が出る可能性を避けたためである。一方、Dublin Core を利用して作られた応用サービスでは、それぞれのシステム構成に応じた表現方法が用いられている。現在、DCMI では、HTML や XML および RDF での記述形式の定義を進めている¹³。

(3) エレメントセットとアプリケーションプロファイル

Simple DC では 15 のエレメントの定義に加えて、「全てのエレメントは省略可能 (Optional) であり、かつ繰り返し可能 (Repeatable) である」と定義している。これに対して図書館での Dublin Core の利用に関するワーキンググループ (DC-Libraries) では、図書館での利用においては、たとえば「タイトルエレメントは必須、主題エレメントは適用可能な場合必須、内容エレメントは記述を推奨」というように提案している。こうした応用毎の適用規則をアプリケーションプロファイル (application profile) と呼んでいる。

ここで、メタデータ規則をエレメントセットとアプリケーションプロファイルに分けて定義しなおしてみよう。エレメントセットはエレメントや限定子など、メタデータとして対象情報資源の性質を記述するための語彙とその意味を定めるものであり、アプリケーションプロファイルはエレメントセットを対象分野に適用する際に定める適用規則である。たとえば、Simple DC の定義は、15 エレメントの名前とその意味の定義に加え、「全てのエレメントは省略可能 (Optional) であり、かつ繰り返し可能 (Repeatable) である」としている。この場合、15 エレメントの定義がエレメントセット、15 エレメントに関して省略可能かつ繰り返し可能と定義している部分がアプリケーションプロファイルということもできる。また、アプリケーションプロファイルは複数のエレメントセットをもとにして定義することも可能である。(図 1 参照。)

実際に Dublin Core を適用する際には資料の種類などに応じた規則の適用方法といった「応用のための指針」も必要とされる。アプリケーションプロファイルはこうした指針をも含めて定義することになる。

(4) 限定子の拡張と Dumb-Down (ダム・ダウン) 原則、構造を持つエレメントの表現

限定子が無秩序に定義されるのを防ぎ、Dublin Core としての相互利用性を保つため、DCMI では限定子の導入に関する基本原則を決めている。これは Dumb-Down 原則と呼ばれるもので、「限定子付きの表現から限定子を取り除いても、当該エレメントとその値の間で矛盾が生じてはならない」というものである。たとえば、「有効になった日付 (Date Valid) は 3 月 20 日である」(エレメント: 日付 (Date), 限定子: 有効になった (Valid), 値: 3 月 20 日) から限定子「有効になった (Valid)」を取り除き、「日付 (Date) は 3 月 20 日である」としても、矛盾は生じない。一方、構造を持つエレメントの場合、たとえば、「作成者の名前は杉本重雄、所属は筑波大学、連絡先は sugimoto@slis.tsukuba.ac.jp」(エレメント: 作成者 (Creator), 限定子: 「名前」、「所属」および「連絡先」、値: 杉本重雄、筑波大学、sugimoto@slis.tsukuba.ac.jp) の場合、限定子を取り除くと「作成者は杉本重雄、筑波大学、sugimoto@slis.tsukuba.ac.jp」となり、筑波大学やメールアドレスが作成者になってしまい矛盾する。この例から、先に述べた値の構造を表す限定子が、一般的には Dumb-down 可能でないことがわかる。

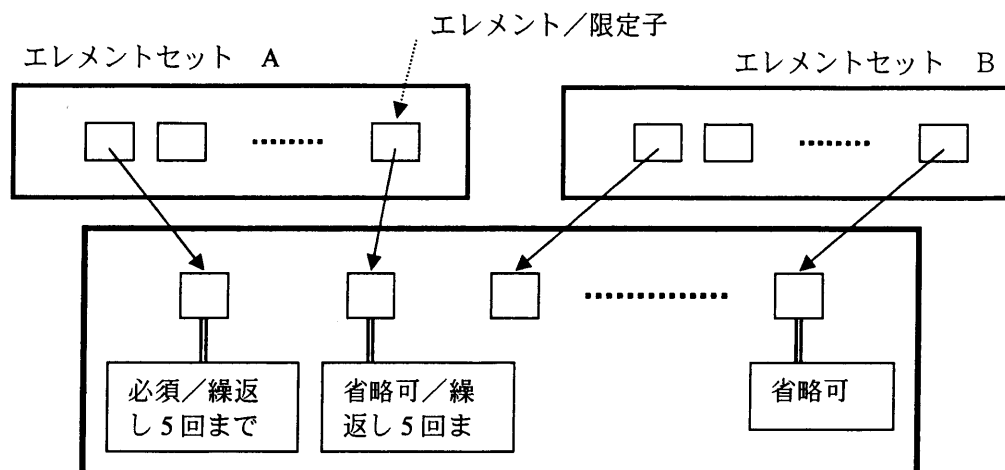
上の例のように構造を持つ値を表現する場合、その値の表現のために必要な属性の集合を別途定義すればよい。たとえば、上の例の場合であれば、別に定義した「名前」と「所属」、「連絡先」を用いて人を表す値を書けばよい。これを RDF のモデルを用いて図示すると図 2 のようになる。また、名前と所属は必須、連絡先は省略可能といった性質はアプリケーションプロファイルにとらえればよい。

3.2.3 Dublin Core の開発と維持

(1) Dublin Core Metadata Initiative (DCMI)

Dublin Core の開発は、OCLC による支援の下に草の根参加者によって進められてきた。Dublin Core の開発の初期の段階では OCLC や NSF 他の助成の下で進められてきたが、Simple DC の合意が得られ標準化を進める必要が出てきたころから DCMES の開発と維持の活動を行う組織を明確にする必要が出てきた。そのため、Dublin Core 開発を支えてきた OCLC がホストする形で DCMI と呼ばれる組織が 1999 年ごろから構成されてきた。その後 Dublin Core の利用が様々な分野で広がるにつれ、新しいエレメントや限定子の導入をも含めた DCMES の維持、地域ごとのコミュニティによる Dublin Core の Localization などいろいろな活動の必要性がでてきた。こうした将来の活動に向け、現在 DCMI の再編成が進められている。

¹³ DCMI のホームページに推奨の記述形式に関する文書 (<http://dublincore.org/documents/>) がある。



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1 エlementセットとアプリケーションプロファイル

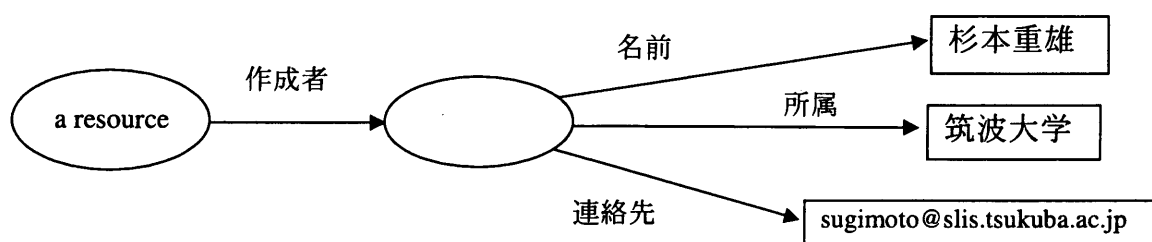


図2 構造を持つ値のモデル

(2) 新しいElementや限定子の導入 – DCMES の維持管理

実際の応用においては、必要に応じたElementや限定子の導入がなされる一方、新しい要素のDCMIとしての標準化も重要な話題である。たとえば、教育分野での利用に関して、どのような利用者に適した資源であるのかを記述するための audience Elementやそれに関連する限定子が推奨記述要素として認められている。政府情報や図書館への適用に関しても同様に議論が進められている。なお、新しい記述要素の推奨や DCMES との整合性の判断は Usage Board と呼ばれる委員会が行っている[47]。Usage Board は各ワーキンググループ、あるいは地域の組織からの新しい記述要素(たとえば、Elementや限定子)の提案を受け、それが DCMES のモデルに適合しているかどうか、グローバルなコミュニティでの利用にとって有用であるかを判断する。

Usage Board ではElementや限定子などに対し、recommended, conforming, registered といった認定を与えている。recommended はグローバルなコミュニティでの利用を推奨するもので、conforming

はElementや限定子が DCMI の基本原則に反しないことを認定したものあることを表す。registered は、LCSH や DDC のように他の責任ある組織によって決められた語彙や表現形式などを表す語として登録されたものであることを意味する。

(3) 分野や地域によるコミュニティと我が国での状況

これまでの開発においては DCMI の場での標準作りと応用毎の DCMES の適用が行われてきたが、今後は政府情報や教育情報といった応用分野の要求に基づく開発、地域ごとの要求、特に英語圏以外の要求に基づく開発が進むと考えられる。また、DCMES の各国語への翻訳、各国での標準の維持なども求められる。こうした活動を進めるには分野や地域ごとのコミュニティを作り上げる必要がある。

我が国では、大学図書館を中心に Dublin Core の利用が広まってきており、日本語で書かれた情報資源に対するメタデータの記述方法に関する知見も蓄積されてきている。また、Dublin Core の国際会議(DC-2001)として東京の国立情報学研究所で開催した。しかしながら、現時点では JIS による標準化はまだ行われていない。日本語あるいは日本の情報

資源のためのエレメントや限定子などに関する議論や意見交換を行うための組織作りが求められている。

3.3 メタデータ規則(スキーマ)の利用

3.3.1 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 3 参照)では、対象を work, expression, manifestation, item の 4 段階でとらえている[48]。電子商取引用の indecs の場合、人(people)、物(stuff)、取引(deal)が記述対象となる[49]。Lagoze 他による ABC metadata model では、作成や更新、変換といったイベントを重要な要素としてとらえている[50]。

Dublin Core は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 のエレメント、限定子、アプリケーションプロファイルを整理しなおすと、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(基本エレメントあるいは詳細化されたエレメントによって表される。)
- (2) 情報資源が持つ属性の値の記述形式(コード化形式限定子によって表される)
- (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(アプリケーションプロファイルによって表す。)
- (4) 具体的な記述形式(保存のための形式とデータ交換のための形式。)

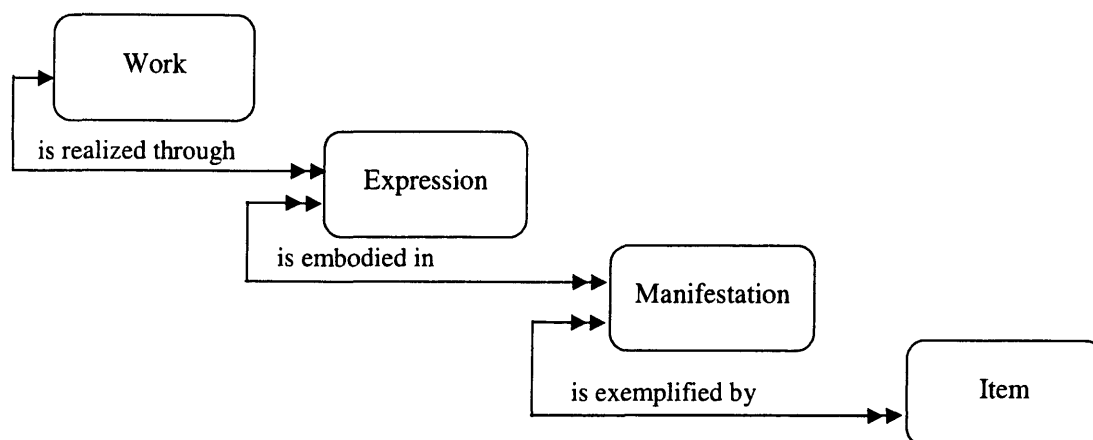
このようにメタデータ規則を整理することで、記述要素の意味(名前とその役割の定義)と構文的性質を分離して定義できる。複数のメタデータ規則(エレメントセット)から必要なエレメントを選んで応用にあったエレメントセットを作り上げることができる。また、エレメント単位での意味的な関係付けができれば、異なる応用間でのメタデータの相互利用が行いやすくなるといった利点を持っている。

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。横断的な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り、そのため、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(Crosswalk)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エレメント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[51]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)が中心的な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。

先に述べたように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。図書館情報大学(筑波大学)では DCMES を中心として多言語によるメタデータレジストリを実現している[52]。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いる[53]。メタデータレジストリは、単にエレメント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、エレメントや限定子に対し、ネットワーク上で一意に定まる識別子を与えることでエレメントなどの意味定義の基底を与えること、エレメント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、新しいエレメント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。



単純矢印は対応する実体がひとつ、二重矢印は対応する実体が複数あることを意味する。両端が二重矢印の場合多対多、一方が二重矢印の場合は1対多の関係を表す。

図3 IFLA FRBRにおける知的内容を表す実体（Group 1 Entity）間の関係

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。ここでは、この3点に関して簡単に述べる。

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、コストが高くつき、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ある。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

4.2.1 デジタルコンテンツの保存に関する基本概念

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、もとのままの状態で保存することにこだわらず、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。後述する OAIS 参照モデルの文書の中でも保存期間に関しては「それなりの長さ」と定義されている。

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。

4.2.2 デジタルコンテンツ保存のシステムモデルー OAIS 参照モデル

OAIS 参照モデルでは、図4に示すように、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 5 に示すように内容情報 (Content Information)、保存記述情報 (Preservation Description Information, PDI)、およびパッケージ化情報 (Packaging Information) からできている。また、情報パッケージに関する記述 (Descriptive Information about Package) も必要とされる。この 4 要素のうち内容情報以外はメタデータと言える。PDI は保存対象コンテンツに関する情報を持つのにに対し、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) : コンテンツの出处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。

図 6 は OAIS 参照モデルの機能実体の関係を表す。SIP, DIP, AIP はそれぞれ寄託情報パッケージ (Submission Information Package)、配布情報パッケージ (Dissemination Information Package) およびアーカイブ情報パッケージ (Archival Information Package) を意味し、コンテンツの蓄積のためのアーカイブへの寄託、利用者からの要求に応じたコンテンツの配布、アーカイブ内での保存のための情報パッケージを意味する。データ管理機能 (Data Management) は、システムに寄託された情報パッケージの登録、利用等の管理を行う。保存用の情報パッケージはアーカイブ記憶 (Archival Storage) に蓄積保存される。利用者からコンテンツのアクセス要求があった場合には、管理機能を用いて検索、利用条件判定などを行い、配布パッケージとして利用者に提供する。管理機能 (Administration) と保存プランニング機能 (Preservation Planning) は、それぞれ保存システム全体の統括管理、状況のモニタリングと蓄積されたコンテンツ保存のためのレポートを行う。

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

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 [42][43][44]。ここでは、OCLC と RLG の共同ワーキンググループのレポートを参考にメタデータについて紹介する。

図 5 に示したように、情報パッケージは内容情報と PDI からなっている。内容情報は、ビットストリームないしビットストリームの集合であるコンテンツデータオブジェクト (Content Data Object) と、このビットストリーム (あるいはその集合) をアクセス可能でかつ意味のある知識の表現に変換するための表現情報 (Representation Information) からなる。表現情報は、コンテンツデータオブジェクト記述 (Content Data Object Description) と環境記述 (Environment Description) からなる。さらに、環境記述をソフトウェア環境とハードウェア環境に分け、ソフトウェア環境は再生用プログラム (Rendering Programs) とオペレーティングシステム、ハードウェア環境は計算リソース (Computational Resources)、記憶装置 (Storage) および周辺装置 (Peripherals) に分けて、メタデータの記述要素 (メタデータエレメント) を定義している。PDI については先に示した 4 つのカテゴリに関してそれぞれメタデータエレメントを決めている。(付録表 1 と 2 に OCLC/RLG のワーキンググループによる主な推奨メタデータエレメントを示す [44].)

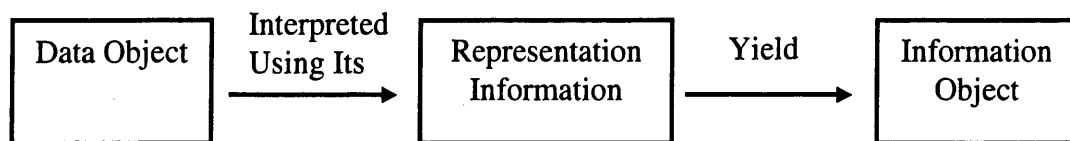


図4 Data Object, Representation Information, Information Object 間の関係

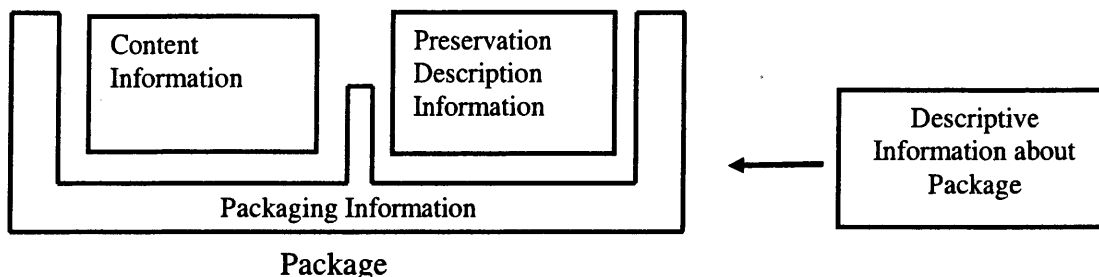


図5 Information Package の概念構造

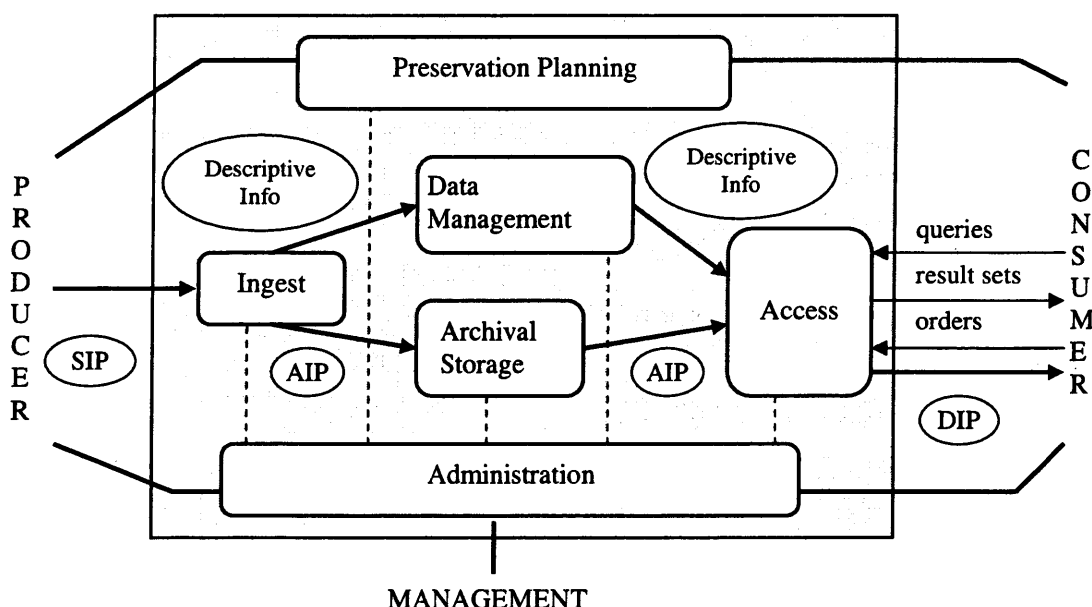


図6 OAIS モデルの機能要素

4.3.2 Metadata Exchange and Transmission Standard (METS)

METS[25] は Encoded Archival Description (EAD)[24]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。

- **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を表す。たとえば、セクションに分かれた録音データの場合、全体とセクションとの関係を記述する。

- Behavior: オブジェクトが持つ動的性質の記述。

4.3.3 コレクションレベルのメタデータ

デジタルライブラリでは資料の集合としてのコレクションを管理の対象とすることがあるため、コレクションの記述のためのメタデータが提案されている。コレクションも保存対象を定義する上で重要であるため、オブジェクト単位の保存のためのメタデータ記述だけではなく、コレクションを対象とした記述も必要とされる。たとえば、カリフォルニア大学バークレー校の場合[54]、印刷物等の従来のコレクションに関して、

- Comprehensive: できる限り網羅的
- Research: 研究および大学院博士で必要とするレベル
- Study: 学部および大学院での教育で必要とするレベル
- Basic: 主題領域を知る上で基本的な資料のレベル
- Minimal: ごく限られた資料だけを収集しているレベル

というコレクションのレベルを定義している。デジタルライブラリの場合、これとは別に、

- Archived: 長期に保存している
- Served: 図書館のサーバに蓄積してサービスをしている
- Mirrored: ミラーサイトを運営して、サービスしている
- Linked: 外部のサーバへリンクしている

といったサービスの違いを明示している。筆者等は、従来のコレクションレベルの記述とデジタルコレクションのレベルの記述を組み合わせ、コレクションレベルでのメタデータの記述の深さを決めるためのモデルを提案した[55]。

4.3.4 書誌記述のための基本モデル - Functional Requirements for Bibliographic Records

図 3 に示した IFLA の FRBR モデルでは、Work は表現方法を問わない知的内容、Expression は何らかの表現方法によって Work を表したものの、Manifestation は Expression を何らかの媒体上に実現したもの、Item は Manifestation の個々のものを意味する。たとえば、「蜘蛛の糸」という小説を仮定すると、作者の考えたストーリーとしての小説は Work であり、それを具体的に表現したもの(小説の場合はテキスト中心に表現されたもの)が Expression である。同じ Work であっても、たとえば大人向けと子供向けといったように異なる Expression を持つ場合がある。たとえば、オンライン書店で「蜘蛛の糸」で検索し、「蜘蛛の糸、ポプラ社文庫－日本の名作文庫(ISBN 4-591-00906-8)」という本を見つけ、それを購

入したとすると、検索結果として示されたものはひとつの Manifestation であり、購入した 1 冊がひとつの Item である。Manifestation はそれから作り出された Item の総称にとらえても構わない。

ネットワークコンテンツの場合、コンテンツとしてのビット列が Expression から Item のいずれになるのか明確ではない。たとえば、コンテンツを利用環境毎にインストールして利用する場合、Item、すなわちコンテンツの実体(ファイル、あるいはビット列)は置かれている場所に依存する個別の実体であると考えられるので、ひとつの Manifestation からいくつもの(異なる)Item が作られるととらえることができる。一方、一般の HTML 文書や単一のイメージデータの場合、Manifestation と Item の間の区別はつけづらい。また、スタイルシートによって表現形式を変化させることのできる XML 文書の場合、XML 文書を Expression、スタイルシートを適用して作られる表示用の文書、たとえば、HTML 文書を Manifestation ととらえることができる。時とともに内容が変わる組織や個人のホームページの場合、Work と Expression が変化し、それに伴って Manifestation も変化するが Manifestation に与えられる識別子(URL)が変化しないという場合もある。

5. Interoperability

Interoperability(相互運用性、相互利用性)は DL における重要なトピックのひとつである。1997 年から 98 年にかけて進められた EU と NSF の共同助成による DL の研究戦略を議論するワーキンググループにおいても interoperability は主要なトピックとして議論がなされた[56][57]。

Dublin Core では Semantic Interoperability ということばキーワードとして用いられてきた。直訳すると意味的相互運用性である。これは「特定の表現形式に依存することなくメタデータを記述するエレメントの意味(だけ)を決め、意味的な相互利用性を保とう」というものである。Dublin Core では、HTML による基本的な記述形式を除いては具体的な表現形式に関する議論はほとんど行われてこなかった。表現形式に関する議論を避けてきた背景には、意味と表現形式の議論は必ずしもうまく切り分けられないため、エレメント定義を進める過程において表現形式に関する議論に引っ張られることを避ける必要があったこと、具体的な表現形式はシステムの実現方法に負うところが大きいことがある。また、エレメントの意味定義に関する合意が形式定義のために損なわれることを避けたためであるとも言える。

Dublin Core は、コミュニティ毎の要求に合ったメタデータ規則のデザインを許すとともにコミュニティをまたがったメタデータの相互利用性を失わないことを

重視している。したがって、コミュニティ毎にエレメントや限定子を導入することを認めるとともに、Dumb-down 原則で相互利用性のための原則を決め、組織としても Usage Board をおいている。Dumb-down 原則のほかに、エレメントセットとアプリケーションプロファイルの分離、およびメタデータスキーマレジストリといったトピックも interoperability の観点からは重要である。エレメントセットとアプリケーションプロファイルを別々に考えることは、エレメントおよび限定子の意味定義と、応用に適用する際の記述上の規約を分離したととらえることができる。両者を分離することで Crosswalk の作成や Dumb-down のための基礎となるメタデータ規則の定義がより明確にできる。メタデータスキーマレジストリはメタデータの interoperability のために重要な役割を持っている。Dublin Core のレジストリの場合で考えると、namespace の管理、各国語訳と英語で記されたオリジナルの記述との間の管理、継続的な開発に伴う版の更新といった点でレジストリの果たす役割は大きい。

Margaret Hedstrom 教授によると、資料の長期保存は temporal interoperability と呼ぶべき、ある種の interoperability である。これは同時期の技術の上に実現する interoperability に比べて、背景の異なる技術の上に作るものなので、格段に実現の難しいものであると言える。Hedstrom 教授は「マイクロフィルムによる保存は原資料の機能を全て保存しているものではなく、それと同じことがデジタル資料にも当てはまってもよいのではないか。作られたときと全く同じ状態で、技術とともに保存しなければならないものばかりとは限らない。」と述べている¹⁴。デジタル資料の保存のために、時を越えてどのような精度の interoperability を求めるかが現実的な課題であることが理解できる。資料保存というと、1 次コンテンツをいかに保存するかにのみ注目が向きがちであるが、利用された技術、コンテンツの保存履歴などメタデータとして記録し、保存する必要があることは言うまでもない。

メタデータのコミュニティからは ontology ということがよく聞えてくる。もともと図書館分野では統制語彙が広く用いられてきたので、それ自身は新しいひびきはあまりない。しかしながら、これまでの統制語彙(あるいはそのために作られたシソーラスや辞書)が、いわば閉じた世界での利用を前提としていたのに対し、ネットワーク上では異なる語彙の基に作られたデータ(あるいはデータベース)間での interoperability, いわば開かれた世界での interoperability がより重要になる。

現在のインターネットでは、様々なコミュニティによって多様な情報資源が発信され、蓄積されている。言語、文化、分野の違いを越えて相互に情報資源を利用できるようにすること、また時間を越えて情報資源の利用性を保つことの重要性は明らかである。コミュニティ毎の高い自由度を保ちつつ、コミュニティ間の interoperability が求められると言える。しかしながら、これを実現することは必ずしも容易なことではない。従来 interoperability を得るためには詳細に決められたプロトコルに従うことが前提であったと思うが、Dumb-down 原則や temporal interoperability の議論にもあるように、緩やかな相互利用性・相互運用性についても理解を深める必要があるように感じる。

6. おわりに - 考察および今後への課題

国家情報基盤 (National Information Infrastructure), 世界情報基盤 (Global Information Infrastructure) ということばとともに、1990 年代のはじめから半ばにかけて、DL に関する研究開発が大きな期待を持ってはじめられた。一方、DL の研究開発と並行して進んだ WWW の爆発的な広がりとともに、我々の情報環境は急激に変化した。たとえば、現在、筆者自身の研究活動はインターネットおよびインターネット上のいろいろなツール無しには全く成り立たない。DL という分野がインターネットに密接に関連したものであることが理由の一つではあるが、研究者一般についてかなり共通していることであると思う。本稿の執筆にあたって、雑誌や論文の論文、種々の報告書など、ほとんどインターネット経由で得ることができた。その中にはかなり 1992-93 年に作られたものも有り、この分野では「古い」資料も含まれている。また、古い資料を探す際に「リンク切れ」が良く問題になる。そうした際に前述の Wayback Machine でリンク切れの資料を見つけさせたこともある。Wayback Machine ですべて解決ということはできないが、資料を保存して提供するという図書館的機能の有用性を実感できた。

国内の図書館現場での DL サービスはこの 5・6 年で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を感じる。その一方、図書館自身による情報資源の開発と提供が求められていることも確かである。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リファレンスサービスなど

¹⁴ 図書館情報大学で開催した第 20 回デジタル図書館ワークショップでの講演による[62]。

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。

DL 分野はいろいろな要素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。EU・NSF 共同の DL の研究戦略に関するワーキンググループで選ばれた 5 分野(知的財産権, Interoperability, 情報発見, メタデータ, 多言語情報アクセス)は、グループの活動から 4 年ほど経つが、これらが情報技術の観点から見た DL の重要な基盤分野であることは今も変わっていないと思う。DLI-1 と DLI-2 を比較すると、DLI-1 では各プロジェクトの中でいろいろな要素情報技術の研究が見られたのに対し、DLI-2 の方がよりコンテンツ指向であると考えられる。これは特定の応用分野のコンテンツとそれに適した情報技術開発を組み合わせた中小規模のプロジェクトが多いためであろう。また、DLI-2 にはデジタル資料の長期保存やサブジェクトゲートウェイ関連の技術開発など、DLI-1 が行われていた間の WWW の発展と電子的な情報資源の急激な増加によって必要性が明確になったものも含まれている。こうした点が DL 分野の特徴を表しているともいえよう。また、教育・学習(e-Learning, 遠隔学習, 遠隔教育), 電子政府, 電子商取引など隣接分野での DL のための技術や知識の適用も重要な話題であると考えられる。

次に、メタデータ分野の観点から、DL において重要と思われる点を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) メタデータを作り、維持するための技術: 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) 多言語メタデータ: 言うまでもなくインターネット上では多様な言語が利用されており、多言語情報アクセスのための情報技術はメタデータの作成や利用においても重要である。

最後に、図書館での DL 間の協調について述べたい。2.2 節に示したように、図書館での DL サービスには典型的なものがいくつかある。DL サービスの場合、地理的な距離が意味を持たない。そのため、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。言うまでもないが DL サービスの場合、コンテンツを提供するサービスのセンターと利用者がコンテンツにアクセスする場所は近くにある必要はない。したがって、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、いくつものサービス間の協調が求められることになる。この観点からは DL 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。

現在の DL サービスに見られる協調のかたちには様々なものがある。たとえば、CORC や GeNii に見られるような分担共同型のメタデータ作成、Open Archives Initiative に見られるようなメタデータ・ハーベスティングによる緩やかな協調、Resource Discovery Network や Renardus プロジェクトに見られるようなサブジェクトゲートウェイの協調、学位論文の分散デジタルライブラリ構築を進める NDLTD[58]を支える大学等の組織間の協調、Z39.50 のような共通のプロトコルを用いた協調など様々である。

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 の記載されたものは、本稿執筆時点(2003年6月)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.nsd.org/>
- [3] American Memory: Historical Collection for
the National Digital Library,
<http://memory.loc.gov/>
- [4] NSF-EU Collaboration,
<http://www.iei.pi.cnr.it/DELOS/NSF/nsf.htm>
- [5] Internet Public Library, <http://www.ipl.org/>
- [6] Global Reference Network,
<http://www.loc.gov/rr/digiref/history.htm>
- [7] Hedstrom, M., "Digital Preservation:
Problems and Prospects", デジタル図書館,
no.20, 2002.3,
http://www.dl.ulis.ac.jp/DLjournal/No_20/1-hedstrom/1-hedstrom.html
- [8] Calanag, L.M., et al, "A Metadata Approach
to Digital Preservation", DC-2001 論文集,
2001.10,
<http://www.nii.ac.jp/dc2001/proceedings/product/paper-24.pdf>
- [9] 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
- [10] 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
- [11] National Digital Information
Infrastructure and Preservation Program,
<http://www.digitalpreservation.gov/ndiipp/>
- [12] 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
- [13] Open Archives Initiative,
<http://www.openarchives.org/>
- [14] CAMiLEON - Creating Creative
Archiving at Michigan and Leeds:
Emulating the Old on the New,
<http://www.si.umich.edu/CAMiLEON/>
- [15] Information Integrity in Distributed
Digital Libraries,
<http://www.prism.cornell.edu/>
- [16] LOCKSS Web Site,
<http://lockss.stanford.edu/index.html>
- [17] Internet Archive,
<http://www.archive.org/index.html>
- [18] 国立国会図書館インターネット資源選択的蓄積
実験事業, <http://warp.ndl.go.jp/>
- [19] JSTOR-The Scholarly Journal Archive,
<http://www.jstor.org/>
- [20] Resource Discovery Network,
<http://www.rdn.ac.uk/>
- [21] Renardus EU-project Home,
<http://www.renardus.org/>
- [22] GeNii NII 学術コンテンツポータル,
<http://ge.nii.ac.jp/>
- [23] MODS: Metadata Object Description
Schema (official web site),
<http://www.loc.gov/standards/mods/>
- [24] Encoded Archival Description (EAD) -
Official EAD Version 2002 Web Site,
<http://lcweb.loc.gov/ead/>
- [25] METS: Metadata Encoding &
Transmission Standard (official web site),
<http://www.loc.gov/standards/mets/>
- [26] Dublin Core Metadata initiative,
<http://dublincore.org/>
- [27] 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
- [28] ARIADNE - Alliance of Remote
Instructional Authoring and Distribution
Networks for Europe, <http://www.ariadne-eu.org/>
- [29] IMS Global Learning Consortium, Inc.,
<http://www.imsproject.org/>
- [30] 先端学習基盤協議会, <http://www.alic.gr.jp/>
- [31] Greenberg, J. (ed), "Metadata and
Organizing Educational Resources on the
Internet", Haworth Press, 302p., 2000
- [32] The MPEG Home Page,
<http://mpeg.telecomitalia.com/>
- [33] 情報処理学会情報規格調査会 SC29/WG11/
MPEG-7 小委員会編, MPEG-7 Japan Home
Page, <http://www.itscj.ipsj.or.jp/mpeg7/>
- [34] GILS: Global Information Locator Service,
<http://www.gils.net/>,
http://www.access.gpo.gov/su_docs/gils/index.html
- [35] AGLS Australian Government
Information Locator Service,

- http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html
- [36] metadata,
<http://www.fgdc.gov/metadata/metadata.html>
(Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998))
- [37] Alexandria Digital Library Project,
<http://www.alexandria.ucsb.edu/> (ADEPT Metadata Standard,
<http://piru.alexandria.ucsb.edu/metadata/>)
- [38] Digital Object Identifier,
<http://www.doi.org/>
- [39] ONIX Product Information Standard,
<http://www.editeur.org/onix.html>
- [40] Indecs Home Page (Interoperability of Data in a e-Commerce Systems),
<http://www.indecs.org/>
- [41] Platform for Internet Content Selection (PICS), <http://www.w3.org/PICS/>
- [42] Cedars Project, "Cedars Guide to Preservation Metadata", 2002,
<http://www.leeds.ac.uk/cedars/guideto/metadata/>
- [43] Lupovici, C., Masanes, J., "Metadata for Long Term Preservation", NEDLIB Report series, no.2, 2000,
<http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf>
- [44] 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
- [45] Dnavi 国立国会図書館データベースナビゲーションサービス, <http://dnavi.ndl.go.jp/>
- [46] Resource Description Framework,
<http://www.w3.org/RDF/>
- [47] DCMi Usage Board,
<http://www.dublincore.org/usage/>
- [48] IFLA, "Functional Requirements for Bibliographic Records",
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>, 1998
- [49] Rust, G., "The <indecs> metadata framework",
<http://www.indecs.org/pdf/framework.pdf>, 2000.6
- [50] Lagoze, C. and Hunter, J., "The ABC Ontology and Model, Journal of Digital Information", vol.2 issue 2, 2001.11,
<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>
- [51] Semantic Web,
<http://www.w3.org/2001/sw/>
- [52] Nagamori, M. et. al, "A Multilingual Metadata Schema Registry based on RDF Schema", Proceedings of International Conference on Dublin Core and Metadata Applications 2001,
<http://www.nii.ac.jp/dc2001/proceedings/product/paper-31.pdf>
- [53] DCMi Registry Working Group,
<http://www.dublincore.org/groups/registry/>
- [54] Berkeley Digital Library SunSITE, "Collection and Preservation Policy",
<http://sunsite.berkeley.edu/Admin/collection.html>
- [55] Calanag, M.L., Tabata, K., Sugimoto, S., "Linking Collection Management Policy to Metadata for Preservation", Proceedings of DC-2002, pp.35-43, 2002,
<http://www.bncf.net/dc2002/program/ft/paper4.pdf>
- [56] An International Research Agenda for Digital Libraries – Summary Report of the Series of Joint NSF-EU Working Groups on Future Directions for Digital Libraries Research,
<http://www.iei.pi.cnr.it/DELOS/NSF/Brussrep.htm>, 1998.10
- [57] Birmingham W. et al., "EU-NSF Digital Library Working Group on Interoperability between Digital Libraries",
<http://www.iei.pi.cnr.it/DELOS/NSF/interop.htm>, 1998.10
- [58] NDLTD: Networked Digital Library of Thesis and Dissertations,
<http://www.ndltd.org/>
- [59] 杉本重雄, "電子図書館－概要と課題－", 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
<http://www.tulips.tsukuba.ac.jp/pub/dlsympo/tutorial.pdf>
- [60] 杉本重雄, Maria Luisa Calanag, "デジタルコンテンツのアーカイブとメタデータ", 人工知能学会誌, Vol.18, No.3, pp.217-223, 2003.5
- [61] International Federation of Library Associations and Institutions (IFLA) の情報のページ. <http://www.ifla.org/II/index.htm>
Digital Library 一般やメタデータに関して IFLA が豊富なリンクを提供している。メタデータに関しては
<http://www.ifla.org/II/metadata.htm> 参照。
- [62] デジタル図書館ネットワーク,
http://www.dl.ulis.ac.jp/dlw_J.html およびデジタル図書館 (ISSN 1345-9198),
<http://www.dl.ulis.ac.jp/DLjournal/>
図書館情報大学で 1994 年以来開催してきたディ

デジタル図書館ワークショップや国際シンポジウムなどの情報、ならびにデジタル図書館ネットワークで発行してきたオンラインジャーナル。冊子体もある。

- Dublin Core に関する情報のページ,
<http://avalon.ulis.ac.jp/DublinCore/>
- PADI – Preserving Access to Digital Information
(National Approaches),
<http://www.nla.gov.au/padi/>

その他の参考資料

- Fox E. ed, “Sourcebook on Digital Libraries, Report for the National Science Foundation”, 1993.12,
<ftp://fox.cs.vt.edu/pub/DigitalLibrary/>
- D-Lib Magazine, <http://www.dlib.org/>
デジタルライブラリ分野での中心的なオンライン
ジャーナル
- Ariadne, <http://www.ariadne.ac.uk/>
イギリスの UKOLN(UK Office for Library and
Information Networking,
<http://www.ukoln.ac.uk/>)によって発行される情
報学分野の専門家、研究者向けのオンラインジャ
ーナル。UKOLN は高等教育・研究機関におけ
る図書館と学術情報サービスのための技術開発
や知識の提供を行う組織。UKOLN のページに
は eLib や Ariadne ほかへのリンクがある。
- Lesk, M, “Practical Digital Libraries – Books,
Bytes & Bucks”, Morgan Kaufmann, 297p,
1997
- Arms, W.A., “Digital Libraries”, MIT Press,
287p, 2000
- Borgman, C.L., “From Gutenberg to the
Global Information Infrastructure”, MIT
Press, 324p, 2000
- Carpenter, Shaw, Prescott (eds), “Towards
the Digital Library”, British Library, 256p,
1998
- 長尾真, “電子図書館”, 岩波科学ライブラリー15,
125p, 1994
- 田畑孝一, “デジタル図書館”, 勉誠出版, 図書
館・情報メディア双書, 155p, 1999
- 杉本重雄, “デジタル図書館(情報機器論, 田畑
孝一編, 第7章)”, 東京書籍, 新現代図書館学講
座 16, 1998
- 日本図書館情報学会研究委員会編, “電子図書
館”, 勉誠出版, シリーズ図書館情報学のフロンテ
ィア 1, 204p, 2001
- 向山, 吉田, 米田編, “インターネット情報流通技
術”, オーム社, 371p, 2000
- 京都大学電子図書館国際会議編集委員会編,
“2000 年京都電子図書館国際会議: 研究と実際”,
日本図書館協会, 228p, 2001
- 筆者の既発表原稿など,
<http://avalon.ulis.ac.jp/~sugimoto/Articles/>

付録表 1 表現情報のための推奨メタデータエレメント（太字はエレメントのカテゴリ）

エレメント名	エレメントの意味
Content Data Object Description	
Underlying abstract form description	ファイルシステムやデータベースの概念要素など内容データオブジェクト（CDO）の基盤的情報に関する人間向け記述
Structural type	静止画や音声、テキストといった CDO が表すデジタルオブジェクトのクラス
File description	ファイル形式や解像度など、CDO を構成するファイルの技術的仕様
Installation requirements	オブジェクトをインストールするために必要な手続き情報
Size	オブジェクトのサイズ
Access inhibitors	暗号化や透かしなどアクセス制限のための CDO の特性記述
Significant properties	将来の保存において保存ないし維持が必須である CDO の利用のために必要な特性。
Functionality	現時点での表示形式に基づいて、CDO の機能あるいは“Look and Feel”の記述。
Description of rendered content	CDO の内容がどのように表示されるべきか、解釈されるべきかの記述。
Quirks	保存プロセスや手続きに起因して起こった CDO の機能や“Look and Feel”の損失の記述
Documentation*	CDO の表示もしくは解釈に必要ないし有用な文書。（以下の各カテゴリにも同様に Documentation エレメントがある。）
Software Environment – Rendering Programs	
Transformation Process	ファイルの解凍やコンパイルなど CDO のインストールのための変換の記述
Display/Access Application	CDO を表示ないしアクセスすることのできるプログラムの記述
Software Environment – Operating System	
OS name	CDO を利用するためのプログラムを実行するプラットフォームの名前ないしは呼称
OS version	OS name で指示されたオペレーティングシステム(OS)の版
Location	OS name および OS version で指示された OS の working copy を入手可能な場所
Hardware Environment – Computational Resources	
Microprocessor requirements	CDO のソフトウェア環境の動作に必要なマイクロプロセッサ仕様の記述
Memory requirements	CDO のソフトウェア環境の動作に必要な記憶装置資源の記述
Hardware Environment – Storage	
Storage information	CDO のソフトウェア環境・利用プログラムの動作に必要な永続的記憶装置資源の記述
Hardware Environment – Peripherals	
Peripheral requirements	CDO の利用ないし表示のために必要な周辺装置の記述
Hardware Environment – as a whole	
Location	CDO の利用に必要な物理的装置を利用ないし入手できる場所

付録表 2 保存記述情報（PDI）のメタデータエレメント

エレメント名	エレメントの意味
Reference Information	
Archival System Identification	アーカイブシステムの中で内容データオブジェクト(CDO)およびそれに関連付けられたメタデータ（保存情報パッケージ, AIP）を一意に識別するための識別子
Global identification	アーカイブ外システムから CDO とそのメタデータ（AIP）を一意に識別するための識別子
Resource description	リソース発見のための情報であり、（利用可能な場合には）もとなるメタデータから抽出、あるいはリソースへのアクセス機能のためにアーカイブ自身によって作られる。
Context Information	
Reason for creation	CDO が作られた理由に関する情報の記録
Relationships	当該 CDO と他の CDO との間の主要な関係の記録
Manifestation	当該オブジェクトと同一のオブジェクトであって、かつ異なる実体との関係。
Intellectual Content	当該 CDO の知的内容と他の CDO の知的内容との間の関係の記録
Provenance Information	
Origin	当該 CDO が作成されたプロセスの記述
Pre-ingest	メンテナンス、内容変更、保管など、当該 CDO 作成から寄託までの履歴に関わる記述
Ingest	当該 CDO がアーカイブに収録されたプロセスの記述
Archival retention	アーカイブ内での保存中における CDO のメンテナンス、内容の変更、管理などの記述
Rights management	当該 CDO に関わる法的事項の記述
Event*	内容や形式、所蔵など当該 CDO に関わる何らかのイベントの記述。（* 上の前エレメントに共通）
Fixity Information	
Object authentication	当該 CDO とその内容の認証のためにアーカイブの最小要求要件に合致するための十分な情報。たとえば、デジタル署名、透かし、チェックサムなど