

# Digital Libraries へのアプローチ — 米国を中心とする取組み事例に基づく考察

杉本重雄

図書館情報大学

〒305 茨城県つくば市春日1-2

tel: 0298-52-0511 fax: 0298-52-4326 email: sugimoto@ulis.ac.jp

## 概要

本稿では、デジタル図書館に関する取組について、米国を中心とする取組事例に基づき考察する。はじめにデジタル図書館に関する基本的機能、基本的要件を述べ、次に自然科学・技術指向、社会科学指向、人文科学指向、マルチメディア情報といったデジタル図書館のタイプに関して考察し、情報の提供技術、情報の可視化技術、多言語文書技術等、技術的要素について述べる。最後に筆者自身の観点から全体的な考察を示す。

## キーワード

デジタル図書館の要件、デジタル図書館のタイプ、デジタル図書館のための基本技術、デジタル図書館プロジェクト、WWW上の情報組織化

## Approaches to Digital Libraries - Issues on Digital Library Projects and Activities

Shigeo Sugimoto

University of Library and Information Science

1-2, Kasuga, Tsukuba, Ibaraki 305, Japan

phone: +81-298-52-0511 fax: +81-298-52-4326 email: sugimoto@ulis.ac.jp

## abstract

This paper describes a study on several issues of digital library projects and various activities related to DL. It firstly discusses several principal functions of and related issues on DL. Then, it discusses the types of digital libraries, which are natural science and technologies, social science, the humanities and multimedia information. It also discusses technological issues to create DL such as information visualization, multilingual document technologies, and information organization methodology on the Internet. Finally, it shows discussions from the author's viewpoint.

## Keywords

Principal Functions of Digital Libraries, Types of Digital Libraries, Technologies for Digital Libraries, Digital Library Projects, Information Organization on WWW

## 1. はじめに

デジタル図書館 (Digital Library, DL) は将来の世界情報基盤 (GII) 上での重要な応用であると認められ米国を始め世界の各国で研究・開発が進められている。DLは非常に大きな応用システムであり、また開発が進められているDLについても分野毎の様々な特徴を持っている。そのため、DLを実現するために解決しなければならない問題には、知的財産権問題をはじめとする社会制度的な問題以外に、全文データ、マルチメディアデータを指向した大規模分散データベース構築とその上での検索技術、多様なデータと多様な利用者のためのユーザインタフェース技術、マルチメディアデータを取り込み、対話し、提供するための自然言語技術や画像・音声認識技術、電子図書・文書を将来にわたって長期的に、また環境に応じて柔軟に利用できるようにするための電子文書技術、情報基盤に向けた新しい情報組織化技術など幅広い情報技術分野の問題が含まれる。

以下本稿では、筆者が昨年来参加した国際会議、大学等での訪問調査、WWWを利用した調査等から得たDLの研究・開発に関する知見を述べたい。筆者はこれまでに二つの論文 [13][14] でデジタル図書館に関する研究動向等を述べてきた。そのため、筆者自身のこれまでの論文に含まれているものについてはできるだけ重複を避けて述べる。

## 2. デジタル図書館 (Digital Library)

### 2.1 図書館の基本機能からのDLに関する考察

#### [図書館の役割と機能]

根本は図書館学研究者の観点からデジタル図書館について述べた [9]。DLは情報基盤上で利用する図書館であると考えられることができる。従って、DLを実現するには、現在の図書館が持つ機能を情報基盤上に実現しなければならない。図書館には、大学図書館、公共図書館、専門図書館等役割の異なるものがある。また、図書館に所蔵される資料についても冊子体の図書や雑誌ばかりでなく、種々の文書資料、写真や絵画、図面、レコードやCD、ビデオテープ、マイクロフィルム、CD-ROM、オンラインデータベースなど様々である。図書館は色々な側面を持つので、図書館が備えるべき機能を一言で言うことは難しい。本稿では図書館の役割を「情報と利用者を結び付けること」と定義しておく。またここで、「情報」は図書や資料そのものやその在りか、図書・資料に収められたデータ、データ間の関連や価値判断に基づくいわゆる情報、あるいはそれらを結び付けるための知識であるとする。

従来開発されてきた図書館情報システムは、オンライン目録システム (OPAC: Online Public Access Catalog) を中心とするものであり、これに加えて発注・受入などの管理業務のためのシステムが組み合わされている。図書館の利用者はOPACで図書の所在を見つけ、その図書が配架されている書架に行って図書を見つけ、閲覧する。あるいは、冊子体やCD-ROM等で提供される索引・抄録を利用したり、図書館員に尋ねたりして資料を見つけ、閲覧する。図書や資料が他の図書館にしかない場合、資料のコピーや図書館間での貸し出しを依頼することもある。

#### [図書館の所蔵資料と情報資源]

図書館の所蔵物は以下のように大きく分けて1次資料と2次資料に分けることができる。またそれらの媒体も紙、マイクロフィルム、CDをはじめとする種々の電子媒体と様々である。さらに蓄積されている資料は日々変化する。

#### 1次資料

図書、雑誌、新聞、灰色文献、絵本、種々の冊子・資料等

音楽・音響資料，ビデオ資料，絵画，写真，地図等

## 2次資料

目録，索引・抄録，辞書・事典，シソーラス等

非常に粗い言い方をすると，DLはこれら（全て）をデジタル化しネットワークを介して利用者に提供しようとするものである。また，図書館の機能を考える上で図書館員の果たす役割を忘れてはならない。DLにおいては，資料を収集し，目録や索引を作る資料の組織化，利用者からの質問に応じて資料や情報を探しだし提供するレファレンスサービス等図書館員の果たしている知的役割の重要性がより一層増すことは疑えない。また，図書館員自身のみならず利用者が情報アクセス技術を修得することを手助けすることも，より一層要求される機能であると思われる。

DL化することによってより多くの情報の中からより適切な情報を見つけだすことができるようになるのであればDL化することの意味はない。そのため，1次情報のデジタル化に注意が向きがちであるが，DLをうまく機能させるには2次情報をいかに使いやすくするかが重要な課題である。たとえば，WWW上を歩き回っていろいろな情報を見て回ることは容易であり，種々のメタインデックスやディレクトリサービスを提供するサイトがあるのでいろいろなサイトの情報を見つけやすくなってきている。それでもなお提供サイトがわからない所望の情報をすばやく見つけだすことは容易ではない。そのため，システムティックに利用できるよう組織化された2次情報を提供することが重要である。

## 2.2 デジタル図書館によるサービス

DLに関する考察を進める上で，いくつかの基本的な話題をあげたい。

### [所蔵物のデジタル化]

前節で述べたように，図書館の資料は大きく1次資料と2次資料に分けることができる。2次資料の中でも各図書館が作成する目録は既にデジタル化が進んでいるものであり，インターネットを介して利用できるものも少なくない。その他の2次資料については索引・抄録や事典・辞書，シソーラスなど，元々検索を目的として作られ，CDあるいはオンラインデータベースとして利用できるものも少なくない。1次資料に関しては，電子テキストとして作られた小数のもの以外は紙あるいはその他の媒体からデジタル化しなければならない。マイクロフィルムのように機械的にscan-inし，デジタルデータ化することが容易なものもあるが，所蔵物の大部分を占める紙の印刷物はページ毎にscan-inするしかない。また，大きな判の資料，貴重書などscannerに適さないものはカメラで入力する必要がある。音楽やビデオ資料についてはアナログデータをデジタル化し圧縮して蓄積する必要がある。また，書誌的情報に加えて内容そのものに基づく索引付け（たとえば，ビデオデータの1カット毎の索引）が必要である。

デジタル図書館の開発は学術分野を中心として進められている。学術分野といっても自然科学・科学技術，社会科学，人文科学といった分野によって情報へのアクセス方法，要求される情報の形式の違いがあるため，所蔵物のデジタル化といっても一様ではないことに留意しなければならない。例えば，scan-inした資料をページイメージで提供するか，あるいは電子テキスト化して提供するかは資料の性質，コスト等に応じて決められるべき問題である。また，人文科学分野ではテキストそのものが研究対象となるので，OCRで入力しただけの検証されていないテキストは利用価値がないといった問題がある。

### [図書館サービスのデジタル環境への適応]

図書館員は図書館の中で最も知的な情報資源であると言える。したがって，デジタル図書館においても図書館員によるサービスを提供するための環境を整えなければならない。山本[17]によると電子図書館員が果たすべき役割を次のように述べている，

- (1) 環境とソフトの整備：ゲートウェイ
- (2) ネットワーク・リファレンス・サービス
- (3) ネットワーク読み聞かせ
- (4) 情報創造・情報発信

をあげ、さらにリファレンスサービスとして

- (1) 情報源アプローチへの協力：ネットワーク・ナビゲーション・パイロティング
- (2) 検索への協力：コオペラティブ・サーチ
- (3) 利用者の知識を生かす翻訳：コオペラティブ・トランスレーション
- (4) オンライン・ツールの使い方指導と探索結果の説明

をあげている。

デジタル図書館へのアクセスのしやすさも重要な要素であると言える。現在、米国では多くの大学図書館が WWW を介したアクセスパスを設け、図書館の案内や利用方法、コレクションの説明、OPAC インタフェースなどを設けている。今後、こうしたアクセスパスはますます充実していくと思われるが、そこで起こり得る問題に次のようなものがある。

- (1) (WWW に共通の問題として) ネットワーク上での 2 次情報の組織化が難しく情報検索が難しい。
- (2) 資料の識別、資料へのアクセス方法が不安定である。例えば、WWW を使っていると同じ URL でアクセスしたにもかかわらずドキュメントの内容が変わっているということは日常的にある。
- (3) 多国語対応の難しさ。図書館に蓄積されている資料は多様な言語で書かれている。ところが、多様な文字を表すための標準的な方法が確立されていない。そのため、例えば、OPAC ひとつをとっても日本から韓国あるいは台湾、中国の図書館にアクセスしてもいわゆる「文字化け」のために本来のテキストを見ることができない。これに類した問題として、異体字の問題、専門分野に依存する特殊な記号や文字の問題をあげることができる。

[ドキュメントデリバリー]

英国図書館のドキュメントサブライセンサーのように世界中の利用者から依頼された資料をコピーし、郵送することを業務としているところもある。また、UnCover [UnCover] や Ariel [Ariel] のようにインターネットを使って資料を検索し、見つけた資料を送付するものがある。これらはドキュメントデリバリー (Document Delivery) と呼ばれ学術文献の提供のために重要な役割を持っている。これらを一歩進め、閲読までをネットワークを介してつながったワークステーションとサーバ上で行うものが現われ、文献の検索から閲読までの機能を提供する「図書館」が作り出された。少し古くなるが、例えば、カーネギーメロン大学図書館が開発した LIS-II (Library Information System II, Project Mercury) は書誌情報によって検索した文献のページイメージをキャンパス LAN に接続されたワークステーション上で読むことができる。UCSF では AT&T の RightPages の上に Springer-Verlag 社の雑誌を載せた RedSage を開発した [RedSage]。また、Elsevier 社は、TULIP (The University Licensing Program) において大学に対して自然科学分野の雑誌のページイメージ、書誌・抄録データ、および OCR 入力した (Uncorrected な) 本文テキストを提供し、いくつかの大学でそれに基づくキャンパスワイドなシステムが作られた。TULIP の「優等生」であるミシガン大学は、TULIP [TULIP] のために開発したソフトウェアを利用して社会科学分野の雑誌記事を提供するシステム JSTOR の開発を進めている [JSTOR]。ドキュメントデリバリーは学術利用では最も基本的な機能と考えられる。そのため、学術情報センターの NACSIS-ELS をはじめとしてドキュメントデリバリーを基本機能とするデジタル図書館システムが多く見られる。

[知的財産権]

知的財産権はデジタル図書館を実現するうえで避けては通れない問題である。現在までに開発が進められた、あるいは現在進められている多くのシステムの場合、プロジェクト毎に出版社等の了解を得て進めている。CD-ROMは安価であるため、広く普及しているがオンラインデータベースの場合、図書館の限られた予算の中では、高価なためそう簡単には購入できない。従来の図書館では資料の利用は原則的に無料であったが、デジタル図書館上では資料の利用は有料になる可能性がある。その際には、いかに課金するか、いかに支払うかといったことが問題になる。いずれにしても、ネットワーク上での情報流通を進めていくうえで、知的財産権問題は社会制度として解決しなければならない問題である。

### 3. デジタル図書館に関する諸要素

デジタル図書館と一口に言ってもそれぞれが大きなシステムであり、かつ異なる所蔵対象、異なる利用者対象を持っている。そのため、ここではデジタル図書館の研究・開発に関連するいくつかの要素について述べる。

#### 3.1 デジタル図書館のタイプ

デジタル図書館は対象とする分野、利用者によってその構成が大きく異なる。そのため、デジタル図書館がどのような分野の資料を提供するのか、どのような利用者を対象とするのかといった、デジタル図書館のタイプを考えておく必要がある。はじめに大きく分けて、コレクション対象によってDLを文献と非文献に分けることができる。前者は図書、雑誌論文や灰色文献、種々のドキュメントを対象とするものである。後者は写真や地図などの静止イメージ、音楽・音声・音響データ、およびビデオイメージやアニメーション等の動画（マルチメディア）イメージである。最も研究・開発が進んでいる学術情報の中で考えると、前者は自然科学（科学技術）、社会科学、人文科学の各分野に分けることができる。自然科学（科学技術）分野には工学、医学、農学分野にさらに細分化して考えることもできるが、ここでは筆者の能力ゆえに自然科学一般としてまとめておきたい。米国においては医学、農学分野はそれぞれNLM(National Library of Medicine)、NAL(National Agricultural Library)が積極的に情報基盤上での情報流通を進めている。イギリスではFollettレポートに基づき、JISC(Joint Information Systems Committee)[JISC]が中心となって高等教育機関の図書館と情報の環境の整備を進めている。

[自然科学・技術分野]

学術文献を対象とするデジタル図書館プロジェクトが最も多く、中でも自然科学分野を対象とするものが多くある。少し古くなるが、カーネギーメロン大学図書館のProject Mercuryにおいて開発されたLIS-II(Library Information System II)では大容量記憶装置としてCD-ROM ジュークボックスを利用し、雑誌記事のページイメージを格納し、書誌情報検索と抄録による全文検索を利用して文献を検索し、見つけた文献のページイメージを画面上で読むものである。このシステムはCMUのキャンパスLAN上に構築されたものである。

UCSFのRedSageは医学雑誌のページイメージを提供しておりフルテキスト検索が可能である。また、ミシガン大学ではElsevier社から提供された雑誌文献(TULIP)のための索引・書誌情報による検索と全文検索機能を持つ検索システムを開発した[TULIP]。イリノイ大学のDL project(Interspace)はIEEE等の雑誌のSGMLに基づく大規模なデータベースを作ろうとしている[UIUC]。CNRI(Corporation for National Research Initiatives [CNRI])では有力な計算機学科を持つ大学と協力して計算機科学のテクニカルレポートのDL(CSTR)の開発を進めている[CSTR]。(CNRIではD-lib Magazine/D-lib Forum[D-lib]でDL関連の話題を集めた記事、論文を提供している。)

工科系の大学である英国の De Montfort 大学では、大学自身によるもの、ヨーロッパの他の機関と共同で進めているものなどいくつかの DL project を進めている [2]。ELINOR[18] は大学内の教育に利用する図書・資料を中心として蓄積したもので OCR を活用して蓄積した目次情報と資料のページイメージを持つ。利用者は OPAC や目次を利用して所望のページを読むことができる。

学術情報センターの NACSIS-ELS は学会誌の書誌情報に基づく検索とページイメージを提供している [1]。奈良先端科学技術大学院大学の図書館は科学技術分野の大学院大学であるという特徴を活かして、雑誌記事をデジタル化（ページイメージあるいは OCR で読み込んだテキスト）して提供する環境の構築を進めている [4]。

自然科学分野の学術文献 DL に共通にいえることは、学術文献を読むことのできる環境を提供することが第 1 であり、そのためページイメージによる提供、全文テキストによる提供という提供形式はコストや文献が提供される形態等の環境に応じて決めれば良いという点である。そのため、TULIP において見られるように、OCR で入力したテキストは（程度にもよるであろうが）誤りを含んでいても、記事を見つけた後ページイメージにアクセスすれば内容を理解することができるので、利用者が必要な情報を得るためには支障がない。また、自然科学分野では図表や写真が重要である。（特に）ページイメージを提供する場合には、画面サイズの制約から、紙に印刷された図表や写真の精度を十分には反映できないことがある。そのため、高精度のイメージを別に用意するなど図表・写真を見せるための工夫が必要である。

#### [社会科学分野]

筆者自身は社会科学分野のプロジェクトについて多くは知らない。ミシガン大学では社会科学分野の学術雑誌（経済学関係）を提供する JSTOR を開発している [JSTOR]。このシステム自身は TULIP で開発したシステムを別の分野の雑誌に適応したものである。WWW の Viewer(Mosaic, Netscape) からアクセスでき、雑誌の巻・号からのアクセス、全文検索によるアクセスを備えている。また、雑誌記事を読むと同時に社会科学関連のデータベースへのアクセスも WWW Viewer から可能であるので記事を調べながらデータベースから統計データを取り寄せるといったことが容易にできる。

#### [人文科学分野]

ミシガン大学の Humanities Text Initiatives(HTI)、ヴァージニア大学の Electronic Text Center(ETC) 他では人文科学分野の研究者のための電子テキストを蓄積している [HTI][ETC]。人文科学分野の場合、テキストそのものが研究対象となり得るので、正確に電子テキスト化されたテキストそのものがようになる。そのため、自然科学分野の場合とは異なり、全文テキストを持つことが必要であり、かつテキストは正確でなければならない。さらに、古文書を多く含むこと、多くの言語の文献を含むことといった特徴のため文字コードセットをいかにするかといった点も考慮しなければならない。たとえば、ETC、HTI では Unicode も利用している。また、ヴァージニア大学の IATH(Initiatives for Advanced Technology in the Humanities) の多言語のエディタ Babble（異なる言語の文書を並べて見ることができ）も Unicode を利用している。

電子化したテキストを将来にわたって長期的に利用するには、印刷形式に依存しない形式で文書の構造を正確に記述することができなければならない。そのため、SGML が用いられている。また、電子テキストの共有を進めるにはメタデータの共有が必要であり、TEI(Text Encoding Initiative)[TEI] のもとに DTD の共有を進めている。また、プリンストン大学の CETH(Center for Electronic Texts in the Humanities) が提供している電子テキストセンターのディレクトリのページで、多数の電子テキストセンター、および電子テキストプロジェクトを見つけることができる [CETH]。

人文科学分野では古書・貴重書に直接触れることが重要であるといわれる。その観点からはデジタル化は無力であるが、一方、古書・貴重書を十分な精度のイメージデータとしてデジタル化することによって世界中どこからでも貴重な文献にアクセスできることに加えて、画像処理の手法を利用した文献の研究に利用することができる。

### [イメージ・マルチメディア]

テキスト検索の場合、検索者が発する質問も書誌情報、抄録、全文と同様にコード化されたテキストであるのに対し、イメージ検索の場合、検索質問に同質のイメージ（静止画、音、動画）を与えることは難しい。そのためイメージ情報の認識技術、音声認識技術、自然言語処理技術などに基づくシステムの開発が進められている。適用分野としては、航空写真や宇宙写真、動植物の写真やイラストなど地球環境情報のライブラリ、K-12に代表される教育分野のためのビデオ情報のライブラリが代表であろう。

航空写真のライブラリ、地図のライブラリ等静止画イメージを対象とするものがある。NSFのDLプロジェクトの一つであるUCSBを中心とするAlexandria Digital Libraryは地図のライブラリのためにメタデータの定義を進めている[12][UCSB]。UCBのEnvironmental Digital Libraryはカリフォルニア州が持つ大量の環境データのDLの構築を進めており、そこには航空写真等のイメージ情報が含まれている[UCB]。また、Pittsburgh大学のPromenadeではNALと協力して植物のイメージを含むデータベースを開発した[11]。電子技術総合研究所でもイメージ情報を利用した感性情報データベースの研究を進めている[6]。

フランス国立図書館ではAudio-VideoデータのDLを計画しており、Audio-Videoデータの索引付けの自動化のためのツールの研究を行っている。NSFのDLプロジェクトの一つであるCMUのInfermedia Digital Video Libraryでは放送局と協力し、ビデオソフトのライブラリの開発を進めている。これは自然言語処理や画像認識技術を用いてビデオ内容を組織化・蓄積し、検索を行うものである[16][CMU]。

## 3.2 FullText vs. PageImage

DLを実現する上で、図書や記事をページイメージで提供するのか、(SGML化した)全文テキストとして提供するのは大きな問題である。雑誌を例にとりて考えてみると、提供すべきデータには以下のものがある。

### \* 書誌情報に基づく検索のための情報

巻・号毎

表紙 (のリスト)

目次 (のリスト)

索引

文献毎

タイトル

著者

巻・号・ページ

抄録

キーワード

### \* 全文

本文全文 (Plain Text)

図表・写真

SGML化したテキスト (図表を含む)

### \* ページイメージ

scan-in して作ったイメージ

画面にはいるサイズの精度のものと印字に適した高精度のもの

ページ記述言語によるページイメージ (PostScript)

前述した人文科学分野のように、テキストそのものが研究対象となる場合 (たとえば、特定の単語が何回現われたか、単語がどのように変化していくかといった問題の場合)、正確な全文テキストが要求されることは明らかである。また、もともと電子出版することを意図したものの場合、ページイメージがなく全文データしかないものもある。一方、自然科学分野の場合のように、図表を含めて「記述を読むこと」が要求される場合、十分な精度で表された図表と共に、記述内容を読みやすいことが要求される。従って、テキストデータがあることが望ましいことは言うまでもないが、遡及入力するような場合、入力コスト、利用方法等の trade-off によってページイメージのみを用意するか、ページイメージと全文データの両方を用意するか、あるいは全文データのみを用意するかについて決定すべきである。たとえば、1994年5月に Rutgers 大学で開催されたワークショップ Digital Libraries: Current Issues においてミュンヘン工科大学の R. Bayer は、広域利用の文献データベース OMNIS/Myriad での経験からコストの点から (抄録の全文データ入力には) OCR を用いるべきであり、人手に頼るべきではないと言っていた。ISDL'95 における OCLC の T. Noreault の講演では SGML とページイメージの比較を行い、その中でページ当たりのコストが SGML で \$15、ページイメージで \$0.5 と述べていた [10]。日本語文献の場合、学術情報センターや奈良先端大、IPA による経験の報告を待つ必要があるが、いずれにしても外国 (アルファベット圏) では scan-in の精度は 300dpi で十分としていたが、400dpi (もしくはそれ以上) で入力しなければならない。

### 3.3 User Interface/Visualization

多様な利用者が多様な情報にアクセスすることが前提となる DL において、ユーザインタフェースと情報の可視化技術は非常に重要な役割を持っていると考えられる。筆者自身は文献 [14] においてこれらについて多少述べたので、ここではそれ以後に得た知見を示すことにしたい。

ADL'95 で講演のあった Galaxy システムは統計的手法を利用して、与えられた語に関連する記事・文献を画面上に点として表し可視化するものである。(残念ながら ADL'95 の講演論文集には論文が含まれておらず、詳細はわからない。昨年の UIST'94 では同様に記事を空間的に可視化する同名の Galaxy が発表されているが、両者の関連については不明である。)

UIUC の Interspace project は INSPEC シソーラスの関連項目を見やすく表示するユーザインタフェースの開発を進めている。これは画面上にシソーラスの項目を 3 次元的に表示するもので、画面上で検索のための語を容易に見つけるための支援ツールとして有用である [5]。

筆者等は協調作業支援システムを利用して GUI を持つ OPAC を教えることに関する評価実験をした [3][15]。その結果、図書館員 (teacher) が被験者 (student) のすぐ横にいて教える場合に比べて、協調作業支援ツールを用いたコミュニケーションのために対話時間は延びるものの、利用者による評価は良好であった。これはデジタル図書館の図書館員が協調作業支援ツールを介して遠隔地の利用者とは対話しサービスすることに関する positive な評価データであると考えている。

### 3.4 ネットワーク上でのドキュメントの組織化

WWW 上で情報を探すときに目的の情報を提供しているホームページの URL がわからないと目的の情報を見つけだすことはなかなか難しい。Net-Orienting とでも呼ぶべきトレーニングが必要になるのではないかと感じる。ところが、頼りにすべき URL も意外に頼りにならない。ドキュメントがアップデートされて以前とは異なる内容を持つものになっても同じ URL の下に置かれるため、時間をおいてアクセス



した場合に以前見つけておいた情報を見失ってしまうということがたびたび生じる。印刷された図書・雑誌の場合にはアップデートが容易ではなく「参照先」が安定しているため「参照の記述」が信頼できるのに対し、URLは余り信頼できない（特に時間の経過したものについては）。DL'95ではURI-WG(Note)のL. Masinterによる講演があり Identifier について考える場が設けられていた。たとえば、電話番号は国際的取り決めの下に個々の回線がユニークに識別でき、図書にはISBNがあり個々の図書をユニークに識別できる。URLは「番地」を一意に決めてはくれるが、その場所に置かれている文書は一意には決めてくれない。そのためネット上の個々のドキュメントを識別できないか、という期待を持つが現実には様々な問題があるようである（注）。

図書館の立場からは資料を如何に組織化すれば効率良く蓄積・利用できるかが重要である。利用者あるいは図書館員自身が必要とする資料がいかに効率良く取り出せるかが問題である。その点ではネットワーク上で提供される資料のメタ情報、すなわち情報の（特性に関する）記述を如何に作るかが問題である。たとえば、図書館の目録は、その図書館で所望の資料を見つけそれが何処にあるかを知るための情報であり、その記述内容を決めるものがこの場合におけるメタ情報である。ネット上には様々な情報があるためメタ情報を決めることは難しい問題であると思われる。OCLCではNCSAと共同でMetadata Workshopを1995年3月に開催し、そこでthe Dublin Core Metadata Element Set（略してDublin Core）と呼ぶメタデータの原形を提案し、議論をはじめようとしている [OCLC]。

（注：URL(Universal Resource Locator)を含むURI(Universal Resource Identifier)に関する検討はIETF(Internet Engineering Task Force)[IETF]の下でのURI working groupによって検討されてきた。URIにおける議論の対象にはURLのほかURN(Universal Resource Name:識別子)、URC(Universal Resource Characteristics:参照のための特性記述)を検討してきた。IETFのURI-WGのホームページによると単一のWGで議論するには問題が広がり過ぎているという理由でWGをcloseしたとのことである (last update Aug. 23, 1995)[URI].)

### 3.5 多言語環境

WWW上では様々な国のホームページを見つけることができる。自分自身はその国の言語を理解できないことはしかたのないことであるが、各国の言語で記述されたドキュメントを自分の画面に表示できないことには大きな不満がこる。WWWの多国語化については議論がなされていることと思うので、ここではデジタル図書館から見た多言語環境の必要性について述べてみたい。

もともと図書館には様々な国の言語で書かれた図書・資料、あるいは様々な国に関する図書・資料がある。従って、図書館情報システムは生来多国語情報を対象としたシステムである。目録情報だけをとって見ても多国語対応でなければならないことは明らかである。現在のシステムの場合は読み替え規則や外字による対応がなされている。この方法は単体システム、あるいは一様なシステムでは通用するかもしれないが、世界に広がるネット上での情報流通の観点からは好ましいものとはいえない。たとえば、他国のOPACシステムをアクセスする場合、入力の形式はアルファベットによる代替が可能であっても、表示内容が文字化けして読めないということは避けなければならないことであると考えられる。1次情報に関しては、なおさら文字化けを避けなければならない。

文字化けを避けるには、世界で通用するコード系とそれに対応するフォントを用意しなければならない。コード系に関して、筆者自身の観点からは、図書館の蔵書には国内標準を作り上げる能力が十分ではない国の言語や、既に古語となっているため標準を決める国や機関がない言語も多く含まれることを念頭においてコード系を決めなければならないと考える。

我々の研究室ではクライアント側にフォントを用意せず、ドキュメントを表示するために必要なフォントだけをサーバからテキストと共にロードするHTML文書のブラウザの開発を進めている [7][8]。ドキュ

メントが長くなってもそこに含まれる字種はそれほど増えない。従って、ドキュメントに含まれる字種のみをフォントをテキストと一緒に送っても転送データ量はドキュメントの長さに比例して増えることはない。必要なフォントをテキストと一緒に送る方式では、これまでの経験からデータ量はテキストの長さの3ないし5倍程度である。

### 3.6 その他 (Miscellaneous)

従来図書館の利用は基本的に無料であった。たとえば大学図書館の場合、学外からの利用者であっても無料で図書・資料を利用できる。一方、先に述べたドキュメントデリバリーや文献複写サービスは有償のサービスとして定着している。多くの資料が電子化されて提供されたとき、何が無料で何が有料なのか、どのようにして利用料を支払うのか、閲覧できる利用者（の資格）をどのように決めるのかといった問題がある。また、いわゆる Electronic Commerce が発展すれば、課金・支払のシステムに関してはそのインフラストラクチャの上に DL も乗ることができるのかもしれない。たとえば、UIUC の Interspace project ではどのように課金するか、といった DL の社会的要素に関する研究も進めている。

デジタル図書館を実際に運営していくことになったとき、デジタル図書館の図書館員の（再）教育をいかに行うか、いかに図書・資料を管理してゆくかといった運営上の問題が現われてくると思われる。

## 4. 考察

本章では、自分なりにこの1年間に得た知見に基づいて、現時点での全体的な考察ならびに印象を述べてみたい。

### [DL 研究・開発の分野]

一概に DL といってもいろいろな目的のものがあるので、ひとまとめに述べることは難しい。DL は総合技術であり、大量の情報を扱う応用システムである。そこで、デジタル図書館を構成するために計算機技術に限らず広く情報技術の観点から重要な技術的要素を（当たり前すぎるものも多くあると思うが、敢て）挙げてみたい。

#### 大規模分散データベース技術

高速ネットワークで結ばれた大規模データベースの応用分野

#### 情報組織化・情報検索技術

（マルチメディア情報を含む）資料の電子化と索引付け

全文検索、マルチメディア検索、内容 (Contents-based) 検索

#### 電子文書技術

文書共有・流通のための標準化 (SGML, HyTime)

マルチメディア文書技術

#### メタ情報化（2次情報化）技術

インターネット上の情報のためのメタ情報化技術

情報源の同定技術（たとえば URI）

#### ユーザインタフェース・情報の可視化

文書・資料を探し出し閲覧し、情報を得るための支援技術

利用者の要求を的確に導きだすための技術

多言語文書・多言語情報

各国のデジタル図書館の情報・資料にアクセスし閲覧するための技術

多言語文書の検索技術

知的作業・協調作業の支援環境

ネットワーク上でのレファレンスサービス，検索支援技術

情報のフィードバック

[DL 環境 - ミシガン大学]

各所で様々な DL プロジェクトが進められているが，その中でミシガン大学の取り組みが非常に印象的であった。それはミシガン大学ではいくつもの DL にかかわるプロジェクトが進められており，個々のプロジェクトが特色あるということに加えて，大学（コミュニティ）全体としてのデジタル図書館環境の向上を進めていることが興味深い。ミシガン大学の Digital Library Initiatives のホームページ [DLI] からの抜粋を以下に示す。

- Campus "Gateway to Information Resources"

ミシガン大学の数多くの WWW のページが提供する情報を利用しやすくするため，形式や用語の統一をはかっている。

- Humanities Text Initiative

人文科学分野の研究教育のための電子テキストを蓄積・提供している。

- Dissertation Express

UMI と協力して行っているもので，学位論文を迅速かつ安価に届けるサービス。

- Core Journals

UMI と協力して行っているもので，図書館の OPAC(MIRLYN) と結ばれ，約 600 のコアジャーナルの記事をイメージとしてアクセスできるようにする。

- JSTOR(Journal Storage Project)

社会科学分野の雑誌記事（10 誌）の提供サービス。TULIP で開発したソフトウェアを利用している。WWW からアクセスでき，記事の検索閲覧ができる。

- Social Science Data Sets

WWW を利用して頻繁に利用される社会科学分野のデータベースへのアクセスを可能にする。

- TULIP(The University Licensing Program)

Elsevier 社から提供された雑誌記事の提供サービス。WWW からアクセスでき，記事の検索閲覧ができる。

- NSF Digital Library Project

NSF から助成を受けて行っているデジタル図書館プロジェクト。地球・宇宙分野の情報を対象とし、エージェントモデルに基づく新しい大規模で開放的なシステムの構築を目指している。

- Newsreel Jukebox

Video Encyclopedia of the Twentieth Century をキャンパスネットワークを通して提供する。

- Papyrology

デューク大学、コロンビア大学とも共同して行っているものでパピルスのコレクションを提供する。

- Making of America

メロン財団がコーネル大学と共同して行っているアメリカの社会史に関する文献をデジタル化し広くアクセス可能にするプロジェクト。

- Museum Information

Getty Art History Information Program が主催する Museum Educational Site License Project のテストサイトとして選ばれており、講義における美術館情報の利用や課金方法ライセンスの提供方法などの研究を進める。

- Electronic Reference Shelf

SGML 検索エンジンを利用し、ネットワークを介した reference 用資料へのアクセスを提供する。

- その他

Geographic Information Systems

University Press Collaboration

E-reserves

Academic Publishing Center

[フルテキスト：SGML]

フルテキストを扱っているところは、当然ではあるかもしれないが「どこもかしこも SGML」という感じがする。既存の雑誌記事、論文等を遡及的に SGML 化することは難しいかもしれないが、新たに出版するものについては SGML 化が強く望まれると感じる。それとともに DTD に基づく文書の標準化による文書の共有が必要であることを感じる。

同じようにフルテキストといっても自然科学分野と人文科学分野で利用目的・方法が異なるため、電子テキストの役割が違うということ（当たり前のことではあるが）忘れてはならない。同様に、ページイメージを提供するものは古く、フルテキスト（SGML テキスト、HTML テキスト）を提供するものの方が新しい、ということも言えない。（テキストは英語やローカルな言語以外の言語、すなわち多言語に対して弱い。）

[多言語化の必要性]

外国に行ったとき、自分の機関のホームページを見せようと思っても文字化けしてしまって見せられない、外国のホームページをアクセスしたが文字化けしてしまって表示できない、といったことをしばしば経験する。これは現在の計算機環境が多くの言語の文書を利用するようには作られていないためである。一般的なオフィスにいる利用者自身がいくつもの言語で文書を書く、あるいはホームページを作るということは余りないかもしれない。しかしながら、デジタル図書館という観点から考えると、1次資料には多様な言語で書かれたものがあり、それにとまって検索に利用する2次情報は多言語となる。これらを同じシステム上で同時に利用したいという要求は極めて基本的なものであり、実用的な観点からの解決が望まれる問題である。

## 5. おわりに

Mosaic の出現により WWW を中心とするインターネット上の情報アクセス機能が非常に利用しやすく、また魅力的な情報の提供と利用が可能になった。現在もすさまじい勢いでインターネット上の情報が増えており、インターネットから情報を得ることが一般化した。しかしながら、WWW 上の情報はダイナミックに変化しているので個人が書き留めることのできる範囲の情報（特に URL）はすぐに古くなってしまふ。インターネット上での情報検索（現在は情報発見とでも言うほうが正確かもしれない）のためにいろいろな Meta-index あるいは Directory Service が立ち上げられている。本稿を書くにあたってこうしたサービスを介して見つけた情報も多く利用した。見方を変えると、既に自分自身がインターネット全体を仮想図書館として利用しているようである。

## 参考文献

- [1] Adachi, J. and Hashizume, H., NACSIS Electronic Library System: Its Design and Implementation, Proc. of ISDL'95, pp.36-41, 1995.8
- [2] Adams, R.J., Electronic Library Research at De Montfort University, Proc. of ISDL'95, pp.42-49, 1995.8
- [3] 後藤清記他, LAN 環境における協調作業支援システムとその図書館サービスへの適用, デジタル図書館 [DLJ], No.3, pp.42-61, 1995.3
- [4] Imai, M, et al., Design of a Digital University Library: Mandala Library, Proc. of ISDL'95, pp.119-124, 1995.8
- [5] Johson, E. and Cochrane, P.A., A Hypertextual Interface for a Searcher's Thesaurus, Proc. of DL.95, pp. 77-86, 1995.6
- [6] Kato, T. and Hirai, S., Human Media Technology - Human Centered Approach to Information Infrastructure -, Proc. of ISDL'95, pp.256-262, 1995.8
- [7] 前田亮他, 組み込みフォントを必要としない WWW のための多言語ブラウザ, デジタル図書館 [DLJ], No.4, pp.21-25, 1995.8
- [8] Maeda, A. et al., A Multilingual WWW Browser without Preloaded Fonts, Proc. of ISDL'95, pp.269-270, 1995.8
- [9] 根本彰, Digital Library は図書館かーある図書館研究者のインターネット体験, デジタル図書館 [DLJ], No.2, pp.15-32, 1994.11
- [10] Noreault, T.R. and Crook, M.A., Page Image and SGML: Alternatives for the Digital Library, Proc. of ISDL'95, pp.145-160, 1995.8

- [11] Rasmussen, E. and McLean, S.A., Image Query and Retrieval: a Case Study, Proc. of ISDL'95, pp.5-9, 1995.8
- [12] Smith, T.R., The WWW Prototype of the Alexandria Digital Library, Proc. of ISDL'95, pp.17-27, 1995.8
- [13] 杉本重雄, デジタル図書館へのアプローチ - 米国で開催された Workshop, Conference に参加して, デジタル図書館 [DLJ], No.1, pp.5-22, 1994.8
- [14] 杉本重雄, デジタル図書館へのアプローチ - DL 関連研究分野に関して, デジタル図書館 [DLJ], No.3, pp.3-19, 1995.3
- [15] Sugimoto, A. et al., Enhancing Usability of Network-based Library Information Systems - Experimental Studies on User Interface for OPAC and of a Collaboration Tool for Library Services, Proc. of DL'95, pp.115-122, 1995.6
- [16] Wactlar, H.D. et al., Technical Challenges for the Informedia Digital Video Library, Proc. of ISDL'95, pp.10-16, 1995.8
- [17] 山本毅雄, 電子図書館員の仕事とその道具, デジタル図書館 [DLJ], No.1, pp.29-38, 1994.8
- [18] Zhao, D. and Ramsden, A., The ELINOR Electronic Library, Advances in Digital Libraries (Preliminary Version), pp.195-210, Springer-Verlag, 1995

## URL

[Ariel] Reserch Library Group の Ariel (ドキュメントデリバリー)

<http://www-rlg.stanford.edu/ariel.html>

[CETH] Princeton 大学の Center for Electronic Texts in the Humanities

<http://cethmac.princeton.edu/CETH/elcenter.html>

[CNRI] Corporation of National Research Initiatives

<http://www.cnri.reston.va.us/>

[CSTR] Computer Science Technical Report project at CNRI

<http://www.cnri.reston.va.us/home/cstr.html>

[D-lib] D-lib Magazine/Forum by CNRI

<http://www.cnri.reston.va.us/home/dlib.html>

[ETC] The Electronic Text Center, University of Virginia Library

<http://www.lib.virginia.edu/etext/ETC.html>

[IETF] Internet Engineering Task Force

<http://www.ietf.cnri.reston.va.us/home.html>

[URI] Internet Engineering Task Force - URI

<http://www.ics.uci.edu/pub/ietf/uri/>

[JISC] Joint Information Systems Committee, UK

<http://www.niss.ac.uk/JASPER.html>

[OCLC] D-lib magazine, July, 1995 の記事

<http://www.cnri.reston.va.us/home/dlib/July95/07weibel.html>

[RedSage] USCF の RedSage に関する記事 D-lib magazine, Aug., 1995

<http://www.cnri.reston.va.us/home/dlib/august95/lucier/08lucier>

[TEI] Text Encoding Initiative

<http://www.uic.edu/orgs/tei/>

[UnCover] CARL の UnCover (ドキュメントデリバリー)

<http://carl.org/carl.html>

ミシガン大学

[DLI] Digital Library Initiatives

<http://www.lib.umich.edu/libhome/digitalprojects.html>

[HTI] Humanities Text Initiative

<http://www.hti.umich.edu/>

[JSTOR] JSTOR project

<http://www.lib.umich.edu/libhome/jstorsumm.html>

[TULIP] TULIP

<http://tulipsrvr.engin.umich.edu/tulip/>

NSF の 6 プロジェクト

[CMU] カーネギーメロン大学

<http://fuzine.mt.cs.cmu.edu/im/>

[Stanford] スタンフォード大学

<http://diglib.stanford.edu/>

[UCB] カリフォルニア大学バークレー校

<http://elib.cs.berkeley.edu/>

[UCSB] カリフォルニア大学サンタバーバラ校

<http://alexandria.sdc.ucsb.edu/>

[UIUC] イリノイ大学アーバナ・シャンペイン校

<http://interspace.grainger.uiuc.edu/>

<http://www.grainger.uiuc.edu/dli/>

[UMich] ミシガン大学

<http://http2.sils.umich.edu/UMDL/HomePage.html>

図書館情報大学

[DLW] デジタル図書館ワークショップ

<http://www.dl.ulis.ac.jp/DLworkshop/>

[DLJ] デジタル図書館 (ISSN 1340-7287)

<http://www.dl.ulis.ac.jp/DLjournal/>

[ISDL] 国際シンポジウム 1995

<http://www.dl.ulis.ac.jp/ISDL95/>