5. 電子図書館における情報入力と情報蓄積

本章では電子図書館における一次情報の入力と蓄積に関する事項について詳しく述べる．

5.1 現状での蓄積形態

本学電子図書館ではまず，紙に印刷された一次情報をスキャナで高解像度で入力し画像情報として記憶装置に格納する．入力された高解像度の画像情報をもとに光学的文字認識（OCR）を用いて印刷されている文字情報を計算機により認識させ，ページに記述されている文字情報の抽出を行う．その後，閲覧表示の際に利用する低解像度のページ画像２種類を高解像度画像情報から生成する．以上をまとめると，本学電子図書館が保有する一次情報は

・高解像度画像
・低解像度画像
・文字認識結果

がある．

また，次節で述べる各種の文書形式のうち重要性の高いものに対応するために平成８年度に予算処置を行い，必要なソフトウェアの開発を行った．これらのソフトウェアは平成９年度より利用可能になり，本学電子図書館が対応できる文書の種類が増加し，これまでよりも一層一次情報の入力の効率化が可能になる．対応した文書形式は，Plain Text フォーマット，HTML フォーマット，Postscript フォーマット，PDF フォーマットである．また同時に後に触れるマスキング処理や本構成作業の効率化も同時に行った．

5.2 既存の電子化情報と電子図書館での保存形式

最近では一次情報が電子的に供給される機会が増えてきている．もともと文書の作成がワードプロセッサを用いて行われていることや最近の電子出版のブームから考えると，自然の成り行きともいえよう．電子図書館にとっても一次情報が電子的に供給されることは一次情報の入力作業の軽減にもつながり，歓迎すべきことである．有効に電子的な一次情報を受け入れるためには，その形式（フォーマット）を正しく理解し，蓄積していくための機構が電子図書館には必要不可欠になってくる．残念ながら，現時点では一次情報を表現するための統一的なフォーマットは確立されておらず，個々の一次情報によってさまざまなものが用いられているのが現状である．次に，現在利用されているいくつかの一次情報表現フォーマットをあげ，それぞれのフォーマットの一次情報を電子図書館で利用する際の問題点について検討を行う．

5.2.1 Plain Textフォーマット

最も基本的な一次情報表現のためのフォーマットである．一般に一次情報となる文書にはさまざまな飾り文字（上付文字や下付文字）や図表などが存在する．Plain Textフォーマットではこれら飾り文字や図表を表現することができないため，文書の表現能力に大きな制約がある．このため，特に数式や図表を多用する科学技術文書などを表現することにPlain Textフォーマットを用いることはまれである．このような文書を電子図書館で利用するためには次のような問題点がある．

・いかに閲覧者に対して印刷物のように見せるか

電子図書館が閲覧要求に対して提示する際に，印刷された一次情報との違いをいかにして少なくするかが問題となる．もっとも，電子図書館が従来のような印刷物に近い一次情報の提示法にこだわらなければ大きな問題ではないが，従来型図書館との連続性を考慮した場合は，それなりの工夫が要求される．電子図書館がその所蔵物をどのように閲覧者に提示するかについては電子図書館のユーザインターフェイスの問題であるので，その方法については別の議論となる．

また上の問題点とは逆に，Plain Textフォーマットを用いる場合の利点には次のようなものがある．

・検索用データの作成が容易である

図書館を電子化することの最大の利点の一つに，これまでの図書館では実現できないような全文検索をはじめとする高度な検索機能の実現がある．このような高度な検索機能を実現するためには，それなりのデータベースを構築しなければならない．

以上のような特徴をもつPlain Textフォーマットであるが，これを本学の電子図書館で利用するためには，利用者に対する提示方法の問題を除いては問題点はない．この問題点を解消するために，平成８年度に必要なソフトウェアの開発を行った．

5.2.2 SGML

SGMLは文書を電子化する際にどのような情報を盛り込んでおく必要があるかということで開発されたもので，文章の本体に「タグ」と呼ばれる修飾子が付け加わったものである．この修飾子はユーザが自由に定義することによって，さまざまな文書の構造を定義することができる．SGMLも一種の計算機言語とみなすことも可能である．ただ、あまりにも柔軟にタグの定義ができるため文書間の共通性・互換性の点で問題が生じる場合もありうる．

SGMLフォーマットの特徴は次のとおりである．

・文書の構造を含めた記述能力が高い

先にも述べたように，タグを文書中に挿入することによって，さまざまな情報を付加することができる．しかも，タグの定義はユーザが独自に必要なだけ行えるので，特殊な文書構造を記述することもできる．

・ページの概念がない

SGMLでは，特に紙に印刷された場合のような「ページ」という概念がない．これまでは情報を記録する媒体（紙）の物理的な制約から，適当な量の情報を記述して紙の余白がなくなれば次の新しい紙（あるいはその裏）に続けて情報を記述していく必要があった．しかし電子的な媒体に情報を記述する場合，これまでの紙のような物理的な制約がない（あっても，これまでには考えられないほど巨大で実用上はないに等しい）ので，特に「ページ」に相当する概念を必要としない．SGMLフォーマットではPSフォーマットのように紙という媒体を意識していないため，紙に情報を記述する際に必要なページの概念がない．我々はこれまで紙という媒体にしか情報を記述することができなかったため，「ページの概念がない」ということに対して多少なりとも抵抗感がある．今後，電子媒体が普及していた場合，紙という媒体の物理的な制約から生まれた「ページ」という概念がどのようになるのか，非常に興味深い．

・タグの互換性が低い

ユーザが独自にタグを定義できるため，タグの定義の仕方が統一されない可能性がある．同じ文書の構造を表現するために，定義の方法がいくつか考えられ，またそれを表現するタグの名称も自由にユーザが決定できる．このため，文書の制作者が異なることによって，文書の構造を示すタグの名称が異なり，可読性が損なわれる恐れがある．SGMLでは一つの文書を一つのファイルにせずに，いくつかの部分に分割し，それをマージする機能を持つ．この機能を有効に活用し，C言語で行われているような効率的なファイル分割（ヘッダファイルの利用）を行うことによってタグの定義の問題を避けることができる．このように，タグの定義を一括して行うようにした場合，文書の配布時にタグを定義しているファイルも同時に配布しなければならない．一般ユーザにとってはどれだけのファイルで文書が構成されているのかが不明確であり，混乱を招く恐れがある．

・図表についての定義がなされていない

SGMLはもともと図表を含まない文書の構造を定義することを主眼として開発された．このため，本学電子図書館が主な収集対象としている図表を多用している科学技術文献については，ユーザが独自にタグとそれに対するアクションを定義する必要がある．この部分についてはSGMLを独自に拡張する必要がある．

一時はSGMLが一次情報記述の主流となると予想された時期もあったが，現在の流れは後に触れるPDFが主流となりつつある．ただ，文書の構造を厳格に記述することができるため，研究分野では依然としてよく使用されており，文書解析をともなう研究では必須ともいえる状況である．また，一部出版社からはSGMLにより一次情報が供給されている．

本学の電子図書館にSGMLフォーマットで表現された一次情報を入力するためには，文書構造を記述しているタグを取り除き，本文のみを抽出するフィルタを用意することによって，該当文書を全文データベースに登録することができる．また，ユーザに文書

を提示する場合には，適当な外部プログラムを用いることによって可能となる．外部プログラムの使用は，適当なソフトウェアが存在すればWWWブラウザの機能によって簡単に実現することができる．また，そのためのサーバプログラムの変更はごく簡単な修正で実現できる．

5.2.3 HTML

HTMLとは Hyper Text Mark-up Language の略で，World Wide Web におけるページ記述言語として使用されている．テキストの記述方法は先の SGML フォーマットとほぼ同じ方法で行うため，SGML のサブセットが HTML であるとの解釈もある．しかし HTML はその用途，目的がSGML とは異なり，また，ユーザがタグを定義することはできない．さらに最近の WWW の発展により，HTML は SGML とは互換性のない拡張もなされている．文書に対する構造の記述法は SGML と同様に行う．図表については，別途作成された画像ファイルを指定することにより本文中に取り込む．

HTML フォーマットがもつ特徴は，SGML フォーマットの際にあげたものと共通する点が非常に多い．ただ，HTML フォーマットはすべての WWW ブラウザソフトであれば表示することができるので，ユーザが WWW クライアントソフトウェアを使用している限り、表示に関しては問題は生じない．SGML フォーマットの場合と同様に，HTML フォーマットからタグを除去する適当なフィルタを用いて文字情報を取り出し，全文データベースに登録することができる．

本学電子図書館ではこのような HTML フォーマットにより供給される文書の入力を効率的に行うため，平成８年度に必要なソフトウェアを開発した．このソフトウェアにより効率的な HTML フォーマット文書の入力と利用が可能になる．

5.2.4 Postscript

もともと Postscript（以下 PS と略）はプリンタへの出力情報を抽象化することを目的として設計された，一種の計算機言語である．PSの記述能力は高く，文章を文字列として表現できるばかりでなく，個々の文字列の印刷位置を指定することもできる．いずれの場合も，紙に印刷された結果は人間にとって同一なものとなる．PSフォーマットの特徴をあげると次のようになるであろう．

- PSフォーマットで記述された文書から文書に書かれている文字情報を意味のある文章（言語）で取り出すことが常に成功するという保証はない．これは記述能力が高いことを意味しているが，このために常に検索用のデータを自動的に生成できるという保証ができない．
- PS フォーマットにより記述された文書は，PS に対応したプリンタに送出すれば紙に印刷できるという利点がある．このため，電子図書館で文書をPS フォーマットで保存した場合，ユーザは簡単に紙に印刷することができる．
- PSフォーマットにより記述された文書をワークステーションの画面上で見るためには，PSインタプリタと適当なフォントがシステムにインストールされている必要が

ある．また，一般にPSの処理系はCPUに対して大きな負荷となる．そのため，資源に制約があるクライアントでPSの処理を行えば，処理時間の増大を招く．また，処理結果はシステムにインストールされているフォントの質に依存した品質となる．通常かなりのメモリ資源を必要とする．

本学電子図書館では，所蔵する文献データに対する全文検索機能を提供しているため，PSフォーマットから文書の文字情報を抽出するプログラムが必要になる．またページ情報をどのような形で提示するか，考慮する必要がある．つまりPSフォーマットからページの画像情報を生成しそれを蓄積しておくか，あるいはクライアント側でPSを表示するヘルパーアプリケーションを利用するかである．現在のクライアント計算機の性能の進歩を考慮すると，ヘルパーアプリケーションを利用する方法がサーバの記憶容量などを考慮した場合有利であると考えられる．本学電子図書館では，このようなPSフォーマットの文書を入力するために必要なソフトウェア開発を平成8年度に実施し，平成9年度より一次情報入力に使用する．これにより，従来は紙にいったん印刷してから入力を行っていたPSフォーマットの文書を直接入力できるため，入力作業に必要な時間や人出，紙などの資源の節約ができ，入力作業の進展が期待される．

5.2.5 PDF

PDFはPortable Document Formatの略で，米国Adobe社が提案したフォーマットである．Adobe社ではPDFフォーマットの文書を作成，閲覧するためのソフトウェア（ビューワソフト）としてAcrobatを開発し，閲覧のためのソフトウェアをAcrobat Readerとして WindowsやMacintosh，あるいはUNIXに対応したものを無償配布している（作成ツールは有償である）．PDFフォーマットはPSフォーマットを発展させたものであり，PSプリンタでPDFフォーマットの文書を印刷できることもある．また，Acrobat ReaderにはPDFフォーマットをPSフォーマットに変換する機能も用意されている．PDFフォーマットの特徴は次のとおりである．

・PSフォーマットとの近親性

PDFフォーマットはPSフォーマットを発展させたものであり，また互換性もある程度は考慮されているので，一部のPDFフォーマットの文書はPSフォーマットとまったく同様に取り扱うことができる．また，PSフォーマットからPDFフォーマットを生成するためのツールも存在する．一部出版社から供給されるPDFフォーマットによる論文の一次情報には，このようなツールを利用してPSフォーマットから変換されたものもある．

現時点では，出版社による一次情報の電子的な供給は多くの場合PDFフォーマットで行われる．Adobe社が提供するAcrobatには先に述べたようなPSフォーマットからPDFフォーマットへの変換機能のほかに，文字情報を抽出する機能もある．この機能を用いて，一次情報から全文情報を取り出し，全文データベースに登録することができる．PDFフォーマットで記述された一次情報の閲覧については，現時点ではAdobe社のAcrobat Readerを用いなければならないが，Adobe社ではアプリケーションを無償で配布

しているほか，WWWブラウザにPDFデータを閲覧するための機能拡張をするプラグインも無償配布しているので，これを利用することにより利用者への情報の提示法の問題は解決する．図書館システムにPDFフォーマットのファイルをユーザに送出できるような機能を組み込むだけでよい．この機能は先のPSフォーマットの一次情報の送出法とほぼ同じであり，PSフォーマットに対応することによってPDFフォーマットも同時に対応できる．

本学電子図書館では平成8年度にPDFフォーマットで提供された文書を直接入力するために必要なソフトウェアの開発を行い，平成9年度より利用可能になっている．これにより，今後出版社からの供給される文書データのフォーマットで主流になると思われるPDFフォーマットが円滑に本学電子図書館に入力することが可能となり，内容充実が容易になった．(参考資料5-2参照)

5.2.6 TeXフォーマット

現状では，ワープロフォーマットによって供給される一次情報は，学内から提供されるものを除いては皆無に近い．学内から提供されるものも，そのほとんどはTeXフォーマットである．TeXとはUNIX上で最も広く使用されているワープロであり，もともとは数学の分野での論文作成のために設計された．その本来の目的からも想像できるように数式の記述が比較的に容易であり，またその印刷時の形式は非常に美しくなっている．また，TeXもSGML同様にユーザがマクロや新しい機能を自由に定義できるような文法になっており，一種の計算機言語とみなすことも不可能なことではない．TeXで記述された文書は「コンパイラ」を用いてDVIファイルに変換される．DVIファイルとは，出力機器から独立して文書情報の記述がなされているファイルであり，DVIファイルを印刷に使用する出力機器（プリンタ）が理解する形式に変換して，紙に印刷する．現在ではプリンタとしてポストスクリプトプリンタが一般的になってきているので，DVIファイルはPSフォーマットに変換されることが多い．TeXは本学情報科学研究科のみならず，世界中の情報科学分野の研究者，研究機関で日常的に使用されている．また，一部（例えば Chicago Journal of Information Theory）ではPSフォーマットやDVIフォーマット，TeXフォーマットで論文を提供しているものもある．

TeXフォーマットについては，本学の成果を積極的に学外に提供するという電子図書館の機能を考慮した場合，何らかの方策が求められる．TeXフォーマットはそのままでは内容の改竄が容易に行われる．内容の改竄を防ぐ意味も含めて，本学情報科学研究科ではコンパイルされたDVIファイルをPSフォーマットに変換したものを公開する原則をたてている．本学電子図書館としては，学内から提供されるPSファイルから一次情報の入力を行う．また，純粋なTeXフォーマットによる一次情報が提供された場合は学内に存在するコンパイラなどを用いてPSフォーマットに変換し，その後入力作業を行うことで対応できる．

5.2.7 その他のフォーマット

これまでに，現在広く用いられている文書のフォーマットについて述べた．しかし，文書を表現するフォーマットは以上にあげたものだけではなく，その他にも非常に数多く存在する．例えば，日常文書作成にはワードプロセッサ（以下ワープロと略）が用いられている．ワープロソフトウェアでは，作成された文書に含まれる文字のフォント形式，フォントサイズなど，さまざまな文書の構造情報を保存する必要があり，そのために各々のワープロソフトウェアに適した形式で文書が保存される．これら，各種ワープロのフォーマットも一次情報を表現するフォーマットであり，電子図書館としてはこれらフォーマットも受け入れることができる能力を備えたほうが望ましい．しかし現在世界中で流通しているワープロソフトウェアの種類は膨大なものとなり，単一の電子図書館が世界中で流通しているすべてのワープロソフトウェアが定義するフォーマットに対応することは事実上不可能である．

先のTeXフォーマットのところで述べたように，現在ではワープロソフトウェアで作成した文書を印刷する際，プリンタとして広く一般的にポストスクリプトプリンタが用いられている．このため，各種ワープロソフトで記述された文書はそのソフトが印刷命令を受領した場合，文書をPSフォーマットに変換してプリンタに送出している．この機能をうまく利用し，各ワープロソフトに依存したフォーマットをPSフォーマットに変換することができると考えられる．このため，電子図書館がワープロソフトに依存したフォーマットで記述された一次情報を受け入れた場合には，そのフォーマットを理解できる適当なワープロソフトを用いてPSフォーマットに変換し，それを受け入れ，データベース登録することによって，ユーザの検索，閲覧に供することが可能になる．

5.3 一次情報入力作業の効率化

電子図書館において，もっとも人的な資源を必要とするのが一次情報の入力であろう．本学電子図書館では，一次情報の入力を可能な限り効率的に実施できるように工夫をしている．現状での作業を示すと次のようになる．

1．これから入力を行う紙に印刷された情報について，スキャナに装着されている自動紙送り機構を利用するために可能なものについては書籍の背を切断し，各ページをそれぞれ別々の紙にする．もしも背を切断できない場合，自動ページめくり機構を備えたコピー機を用いて全ページのコピーを行い，各ページ別々の紙になるようにする．

2．ページ情報の入力をグレースケールスキャナを用いて行う．このグレースケールスキャナは自動紙送り機構を備え，紙の両面を自動的に読み取ることができる．ただし，これらの機能は個別の紙に対しての機能であり，製本されたままの書籍に対しては無効である．このため，1に述べたような方法で個別の紙にしておく．

3．もしもカラー情報が含まれていた場合，別途カラー情報を含むページをカラースキャナで読み取る．カラースキャナは先のグレースケールスキャナとは異なり，紙送り機構が装着されていない．このため，紙をスキャナの上に置く作業は手作業に頼ることになる．

4．ページ情報の読み取りは，元の書籍のページの順序には支配されずに行うことができるが，ページ情報読み取り後にページ情報の順序を決定する作業が必要になる．この作業を構成作業と呼んでいる．

5．入力中の書籍について，その二次情報（題名，著者，目次など）の情報を入力する．このとき，例えば他の学外のデータベースなどにすでに登録されており，その情報を利用することができるのであれば，その情報を利用する．

6．構成作業が終了した後，システムが必要とするすべての情報（ページに書かれている文字の認識や閲覧用画像）の生成を行う．（この作業は本構成作業と呼ばれている）また，この文字認識の作業を実施するかしないかはオプションで自由に選択できるようになっている．

通常の印刷された一次情報については以上の様な手順で電子化を行っている．この手順の中で最も時間と人手を要する行程は3のカラー情報の読み取りおよび5の二次情報の入力作業である．特に5の二次情報の入力については，本学電子図書館システムに入力すべき情報を網羅した，学外に利用できる適当な二次情報のデータベースを発見できていないため，自動化を妨げている．この作業に有効な二次情報のデータベースの発見が重要な課題である．また，その種のデータベースが発見された場合，そのデータベースのフォーマットから本学で二次情報入力に利用しているフォーマットへの変換が課題となる．このフォーマット変換については後にその解決法を示す．

これまでは書籍から一次情報を入力する場合について述べてきたが，一部出版社からはページの画像情報とそのOCR結果，書籍の二次情報がCD-ROMにより供給されている．これらの出版社の情報の入力については，上記の1から3までの手順が省略され，次の1’の手順が行われる．

1’出版社から提供されている画像情報と文字情報をしかるべき格納場所に保存する．

このようにすることで，本学で書籍を入力した場合と変わりなく閲覧に供することができる．書籍をスキャンする作業と文字認識が出版社で実施されているので，本学では構成作業と閲覧用の画像情報の生成を行うだけでよい．さらに，供給されている情報には書籍の二次情報もある．供給されている二次情報のフォーマットは出版社が独自に定義したものである．これを本学の電子図書館システムで定義されているフォーマットに変換するためのソフトウェアを開発することにより，先にあげた5の作業を省略することができる．なお，この変換を行うソフトウェアの開発は本学のリサーチアシスタントの制度を利用して，本学情報科学研究科の博士後期課程の学生に作成を依頼した．この出版社の場合のように書籍情報を電子的に供給を受けた場合，当然のことながら一次情報の入力作業が軽減される．

また，カラーページの入力については現在は個別に該当するページを手作業で選び出し，作業を行っている．さらに出版社との契約によって広告などページの一部分をマスキングする必要があるものもあり，これらは本構成作業の前に別途作業が行われる．これらの作業については次に議論する．

5.4 マスキング作業の効率化

出版社との間での書籍を電子図書館で利用するために結んだ契約において，何らかの著作権上の理由により書籍のページの一部写真，広告などをマスキングすることを条件に電子化することになっているものがある．この条件を満たすために，電子化作業の途中で指定されたページの一部分をマスキング処理し，ユーザの目に触れないよう作業を行っている．この作業は，読み取ったページの画像情報に対して，人手により別途マスキング作業を行っている．マスキング作業はMacintosh上のPhotoShopを用いて行っている．本学の電子図書館はUNIXをベースとしたシステムで構築されているが，マスキング作業のみMacintoshで行われている．このため，入力されたページの画像情報は、Columbia Appletalk Package（以下，CAPと略）を用いてMacintoshシステムとの間でファイル共有を実現し，マスキング作業を実施している．CAPはメルボルン大学で開発，維持作業が行われているフリーウェアで，UNIXシステムでMacintoshのためのAppleTalkサービスを実現するものである．CAPの問題点は，フリーウェアであるため，ソフトウェアのバージョンアップが頻繁にかつ予告なく行われる点である．さらに，ソフトウェアの実行速度が遅いことも作業効率の低下を招いている．作業効率を高めるためには，CAPあるいはそれに変わるUNIX上でAppleTalkサービスを実現するソフトウェアの高速化もしくは，UNIX上で動作するPhotoshopを導入するなどの方策が必要である．このため，本学電子図書館では平成８年度に予算処置を行い，一連のマスキング作業をUNIX上で行うために必要なPhotoshopの導入及び必要なソフトウェア開発を実施し，作業効率を高める処置を行った．このソフトウェアは平成９年度より実際の入力作業に利用する．

5.5 カラー情報を含む一次情報入力の高速化

カラー情報を含む一次情報の入力法は前に述べた．その過程で最大の問題となっているのは，カラーページの入力装置に自動紙送り装置がない点である．当初，システム運用開始前には，電子化の許諾が得られた書籍にはそれほど多くのカラーページはないであろうと予想していた．しかし，ある種の雑誌にはカラーページが多数含まれているため，当初の予想をはるかに上回る比率でカラーページの処理が発生することになった．

カラーページのスキャニングはモノトーンの原稿に比べて原理的に３倍の時間がかかる．また自動紙送り機構がないため，ある一定量のページを装置にセットして自動的にスキャニングすることができない．現在のスキャナ市場をみてみると，モノクロスキャナには用意されている両面に対応した自動紙送り機構が，カラースキャナには用意されていないようである．カラー情報を含んだ一次情報の入力作業の効率化のためには，両面に対応した自動紙送り機構を備えたカラースキャナの開発が待たれるところである．また当該装置が開発されるまでの暫定的な処置として，平成８年度にカラースキャナを増設し，カラーページの入力を効率化する処置を行った．

5.6 記憶メディア――マイグレーションシステム

電子図書館では，すべての一次情報を電子化して蓄積するために巨大な記憶装置が必要になる．本学の電子図書館は情報科学分野，バイオサイエンス分野，および物質創成科学分野の３分野のみの学術情報書籍の収集しか行わない．しかし，それでも最終的には10TBを超える一次記憶装置が必要とされ，これをすべてハードディスク装置で構成することは非現実的である．現在，記憶メディアとして実用化されているもののうち一般的なものは，アクセス速度の順に半導体メモリ，磁気ディスク，光磁気ディスク，磁気テープがある．このようなデバイスを用いて超大容量の記憶システムを構築する必要があるわけであるが，記憶システムの構築に許される費用にも限界がある．

理想的にはアクセス速度が最も速い半導体メモリだけでシステムを構築することが望ましいが，構築費用，維持費用を考えたとき10TB以上の半導体メモリによる記憶システムは現実的でない．

次に考えられるのが磁気ディスクのみでシステムを構築することである．ここ十年以上の間で磁気ディスクのビットあたりの単価は考えられないほど低下してきている．それでも，10TB以上の記憶システムを磁気ディスクだけで構築することは費用的に考えても実現性が乏しい．

光磁気ディスクシステムは，メディアを着脱することができ，ジュークボックスを利用することにより装置あたりの記憶容量を大きなものにすることができる．この２，３年前まではビットあたりの単価が磁気ディスクと比較するとかなり安価であったので，費用の点での有利性が大きかった．アクセス時間はジュークボックスシステムであるため，また，光磁気ディスクの原理的な問題からそれほど速くない．最近の磁気ディスクの価格低下のため，構築費用の点やアクセス時間の点からみた魅力は薄れつつあるのが現状である．

磁気テープシステムは最もビットあたりの単価が安く，ジュークボックスを利用することにより大容量の記憶システムを構築することが可能である．アクセス時間はジュークシステムであるためとテープというシーケンシャルデバイスであるため，それほど速いものではない．しかし，ビットあたりの単価が磁気ディスクとは比較できないほど安いため，最も現実的な記憶システムである．またアクセス時間については，さまざまな技術的工夫により，シーケンシャルアクセスデバイスである欠点が克服されつつある．物理的には，安価な磁気テープシステムは耐久性の点で問題がある．実用化されている磁気テープシステムには，8mmビデオテープを利用したものやDATテープを利用したもの，放送用ビデオテープシステムを利用したものなどがある．8mmビデオテープやDATテープは巻き戻しや早送りが頻繁に発生するコンピュータ記憶メディアとしては耐久性に問題がある．これらのテープは通常，数十回のアクセスでテープの寿命が尽きてしまう．放送用ビデオテープシステムは現在放送業界のディジタル化により，放送局で利用されはじめている．放送局での利用では，同一のテープを繰り返し利用する．このため，耐久性は8mmビデオテープシステムなどとは比べ物にならないくらい高く，コンピュータの外部記憶メディア，特に図書館での情報蓄積メディアとして適当である．

電子図書館では，蓄積されている情報に対して利用者からの要求があったときに可能な限り短い時間で情報を提供する必要がある．このような目的と，現実の費用的な問題も考慮にいれると，磁気ディスクと磁気テープを組み合わせた超大容量記憶システムの構築が適当であると考えられる．本学電子図書館は，そのシステムの基本的な設計は数年前から行われていた．その当時は，光磁気ディスクの磁気ディスクに対する優位性が顕著であったので，磁気ディスク，光磁気ディスク，磁気テープの三層構成のシステムが費用的に最も有利であった．

デバイスとして複数の種類のものを採用した場合，その費用によって構成される容量が異なる．磁気ディスクが最も高価であるのでその容量が最も小さく，磁気テープはビットあたりの単価が安価であるので，容量が最も大きい．このように容量の異なるデバイスをまとめ，一つの記憶装置のようにみせる技術が必要である．これがマイグレーションシステムと呼ばれるものである．マイグレーションシステムでは，アクセス頻度の高いものをアクセス速度の速いデバイスに置き，アクセス頻度の低いデータをアクセス速度の遅いデバイスに置く．この過程はユーザにはまったく意識させることなく行われる．これは電子図書館システムにおける用途と一致する考え方である．一般に図書館とくに研究図書館においては，頻繁にアクセスされる書籍とそうでない書籍に二分される傾向にある．つまり新しい文献や，古くても重要な文献に対しては研究者は頻繁にアクセスするが，古い文献にはあまりアクセスをしない．このような図書館が保有するデータの特性とマイグレーションシステムの考え方は一致するものがある．

マイグレーションシステムは汎用機の分野では古くから実用化されていた．たとえば金融機関などにおいて超大容量記憶システムの需要があり，技術開発が進んでいる．一方，UNIXベースのシステムでも商品として一部出荷されているが，UNIXシステムにおける超大容量記憶システムはその需要が始まったばかりである．本学電子図書館を始めとして，今後超大容量記憶システムに対する需要が発生し，その技術開発が一層進むものと思われる．

5.7 ビデオ情報の入力

電子図書館の役割の一つとして，各種文献の収集のみならずオーディオビジュアルデータの収集，蓄積が大きなものとしてある．また，大学図書館の場合にはさらに，学内で発生した各種学術情報（論文などの文献情報のみならず，研究成果のデモンストレーションビデオなど）を学外の研究者に対して発信する責務がある．このような機能を果たすため，本学電子図書館ではビデオ情報の蓄積，配信を実現すべくビデオ情報も電子化して蓄積配信する体制を整えている．

ビデオ情報の電子化については，現在世界各国でその研究，開発，実用化が行われつつある．実際，米国では家庭に対するテレビ放送をディジタル化することが決定され，現在その準備が行われつつあり，わが国においても郵政省がテレビ放送をディジタル化する方針を検討している．このような環境の中，本学電子図書館でもビデオ情報の電子化には積極的に取り組んでいる．

ビデオ情報を電子化する際の規格は，いくつか提案されている．最も早くに規格化されたものにMPEGがある．この規格ではとにかくビデオ情報を計算機システムで取り扱うことを目標に開発されており，画質の面で十分とは言えない点があった．この欠点を克服し，放送局でのビデオ編集に耐えうる画質を提供する規格としてMotion JPEGと呼ばれる規格が開発された．Motion JPEGでは一秒間に30枚送出されるNTSCビデオ規格の画像を画像の連続性を考慮し一枚一枚JPEGと呼ばれる圧縮方法で圧縮し，取り扱うものである．JPEGの規格では圧縮率を画像に応じて設定することができるため，高画質を要求される場合は低い圧縮率で，画質が重視されない場合には高い圧縮率で画像を圧縮することができる．Motion JPEGでは画像の連続性をある程度考慮しているが，より積極的に利用する余地は残されている．この点を改善したのがMPEG2と呼ばれる動画像圧縮形式である．MPEG2では，動画像の性質により圧縮率を変更できる．例えば，スポーツ中継のような動画像では画像間の動きが激しいため低い圧縮率（高いビットレート）にすることにより，高品質な動画像とすることができる．これとは反対に，動きの少ないニュースや講義などの動画像の場合は高い圧縮率（低いビットレート）にすることにより，画質を損なうことなく必要なデータ領域を節約することができる．このMPEG2は各種のディジタルテレビ放送で用いられている圧縮方法であり，世界的な標準技術となりつつある．

本学電子図書館ではこのような技術動向を鑑み，MPEG2を基本としたディジタルビデオ蓄積，配信システムを構築した．このシステムでは，MPEG2によりエンコード（符号化）されたビデオ情報を高速なサーバに保存し，ユーザはクライアントから欲するビデオ情報にアクセスする．ただしビデオ情報のビットレートは4Mbps以上と高いので，伝送速度が10Mbpsのイーサネットでは十分に送れない問題がある．そこで，伝送速度が100MbpsであるFDDIネットワークを用いてビデオ情報の伝送を行うこととした．このため，本学情報科学センターは従来から用いている本学の基幹ネットワークとは別個にビデオ情報配送のためのマルチメディア専用のネットワークを整備し，ビデオ情報が障害なくサーバからクライアントに送付できるようにした．さらに，MPEG2ビデオ情報を

再生するためのハードウェアの整備も行っている．

本学電子図書館では，著作権の許諾が得られたものから順に電子化作業を行っている．電子化しているビデオ情報には学外で制作された制御工学関係の講義ビデオなどがあり，利用者である学生がこの種の自習を行う際に利用されている．

5.8 ビデオ情報の編集

前節ではビデオ情報の入力とその保存形式について述べたが，ビデオ情報を学内で制作するためには編集作業が必要になる．ビデオの編集作業には，通常長い時間と経験が必要とされている．この大きな理由の一つに，ビデオ情報はテープに記録されているため，ランダムアクセスができないことがあげられる．現在，一般的に行われている編集方法は次の通りである．

1．編集すべきビデオ情報を一通り見る．
2．作成するビデオの構成に基づいて，必要なカットを選び出す．
3．選び出されたカットを所望の順序に並び変える．
4．並び変えられた順序に従って，元テープを再生し，それを新しいテープに録画する．

現在のテープを基にしたビデオ編集システムでは，この編集作業を行う上で次のような不都合がある．

1．必要なカットが一本のビデオテープの前半と後半にある場合などはそれらを同時に見ることができない．
2．カットからカットへ移動する際には，テープの巻戻しもしくは早送りが必要になり，時間がかかる．

この不都合による編集作業の効率低下を防ぐためには，多大な編集経験が必要になり，本学のように専門的にビデオ編集を行う人材を確保できない場合にはビデオ情報の制作が大きく制約されてしまう．先にあげた不都合の原因はビデオ情報がテープに記録されており，そのアクセスが線形的に（リニアに）しか行えないことである．もしも，ビデオ情報へのアクセスがランダムに（ノンリニアに）行うことができれば，編集作業における時間と労力の無駄を省くことができる．最近のコンピュータ技術の発展により，ビデオ情報をディジタル化してハードディスクに保存し，その管理を容易に行うことが可能になってきた．このため，ノンリニア編集システムが開発されつつある．

本学電子図書館では，平成８年度にノンリニア編集が可能なビデオ編集システムを導入し，学内で容易にビデオ情報を制作することができるようにした．