第２回JHKオープンセミナー開催記録

デジタル化された人類の記録は次代に残せるのか？

「デジタル情報の長期的な保存とアクセスのために」

国立国会図書館関西館事業部電子図書館課　今野篤（こんのあつし）氏

司会　第２回ＪＨＫオープンセミナーを始めさせていただきます。本日はＪＨＫ会長の金澤勇二よりごあいさつのあと、国立国会図書館関西館の今野先生より「デジタル化された人類の記録は次代に残せるのか？　『デジタル情報の長期的保存とアクセスのために』」というテーマでご講演いただきます。最後に 30 分ほどディスカッションを行いますのでよろしくお願い申し上げます。

金澤　今日はお暑い中お忙しい中、たくさんの方にお集まりいただきまして有難うございます。

情報保存研究会は 2000 年にスタートしました。 はじめは会員を対象とした勉強会を行いましたが、去年から一般の方を対象としたオープンセミナーを開催するようになりました。情報保存研究会は情報保存機関にかかわる企業の集まりで、われわれ企業の技術やサービスを情報保存機関の方に紹介するとともに、あわせてその発展振興に寄与しようと活動しているグループです。活動のメインは今日のオープンセミナーのほか、ダイレクトリーを発行しています。今年の 4 月にはホームページを開きました。ダイレクトリーとホームページには、公文書館などを中心とした情報保存機関に対する情報保存に関するアンケートをお願いしてその結果を載せております。また Q&A として情報保存に関する技術や知識を載せております。このようにＪＨＫの活動は少しずつ充実してきております。

最近デジタル化された個人情報が故意やうっかりで世の中にばらまかれているという事件が毎週のように騒がれております。クラッカーが新しいウイルスをばらまいてそれを防ぐのに困っていたり、自分のパソコンが壊れてハードディスクに入っていた情報がなくなってしまったりという経験を持っている方がいらっしゃると思います。本日はそのようなセキュリティーの問題やデジタル情報の長期保存の問題についてお話いただきます。わざわざ京都からお越しいただきました講師の国立国会図書館関西館の今野さんと沢山お集まりいただきました皆様にお礼を申し上げまして挨拶に代えたいと思います。

司会　金澤会長、ありがとうございました。それでは早速セミナーに入りたいと思います。

本日の講師の先生をご紹介いたします。国立国会図書館関西館事業部電子図書館課にお勤めでいらっしゃいます、今野篤先生です。

「デジタル化された人類の記録は次代に残せるのか？　『デジタル情報の長期的保存とアクセスのために』」というテーマでご講演いただきます。それでは、よろしくお願いいたします。

今野　よろしくお願いいたします。今野と申します。

「デジタル化された人類の記録は次代に残せるのか？　『デジタル情報の長期的保存とアクセスのために』」というタイトルでお話をさせていただきます。

１．現状確認

まず、現状を簡単に確認させていただきます。ボーンデジタルというものがあります。ボーンデジタルとは､最初からデジタル情報､デジタル形態で作成される情報のことです。このようなものが近年非常に増えております。

例として銀塩カメラとデジタルカメラの出荷数量の推移があげられます。デジタルカメラは 2001 年に国内出荷数量が銀塩カメラを上回りました。2002 年には輸出分も含めてデジタルカメラの生産量のほうが多くなりました。DVD ビデオは 2003 年に国内出荷数量が今までのビデオテープレコーダーを超えたそうです。

ボーンデジタルの代表的なものといえるのがウェブです。表層、深層と分けることができます。ウェブの表層というのは静的なページで、HTML 等で書かれているページです。深層というのは逆にデータベースやアクセスするたびに動的に作られるページのことです。ウェブには表層ウェブと深層ウェブ、二つのカテゴリーがあります。

How much information という調査がカリフォルニア大学でされています。この調査では、表層ウェブは 2002 年の時点で 167 テラバイト、深層の方は 91 ペタバイトあるだろうという推測をしています。ペタバイトはテラバイトの 1,000 倍ですから、とんでもない量のウェブ情報があるということになります。

メールも大量に発信されていて、１日 310 億通が送信されているそうです。メールなしの仕事は最近では考えられなくなりましたが、この数は増えていく一方だと思われます。

このような流れを支えているのが技術革新です。このデジタル情報関係の技術革新を表す良い例だと思われるのがムーアの法則で、これは半導体の集積密度が 18 から 24 か月ごとに倍増していっているという経験則です。インテルの創設者の一人のムーア博士が経験則として提唱したものです。コンピュータ以外の技術、例えば自動車などで最高速度が 1,000 倍になったとか燃費が１/1,000 になった・なるということはあり得ません。デジタルならではの技術革新の速度だと思います。このムーアの法則ですが、半導体の微細加工技術の発展を根拠としていて、2010 年ごろには、その微細加工技術が原子レベルにまで到達してしまい頭打ちになるとも言われております。

しかしムーアの法則が仮に頭打ちになって半導体の集積密度がそれほど上がらなくなったとしても、その他にも技術革新が進んでいる領域がいくつもあります。例えばギルダー

則というものがあります。これはネットワークのスピード、帯域幅がムーアの法則の 3 倍で向上するという経験則です。ネットワークのスピードが上がれば上がるほど、重いコンテンツが多く作られていくと思います。昔は Flash を使ったコンテンツはそれほどなかったと思いますし、画像のサイズもそれほど大きくはなかったと思います。昔は、14.4kbps のモデムなどを使っていたのですが、最近はブロードバンドが一般的になり重いコンテンツが増えたという印象です。

また、ハードディスクの容量も急速に増加しています。私が大学のころ、15 年くらい前ですが、40 メガバイトのハードディスクは大容量だったという記憶があります。今では 120 ギガバイトなどあたりまえですね。120 ギガを 40 メガで割ると 3,000 倍です。ハードディスクの容量が 15 年で 3,000 倍増加したということです。

さらに CPU の動作周波数、CPU の性能を表す一つの指標ですが、こちらも急速に伸びてきました。これも 15 年ぐらい前は 10 メガヘルツ程度だったと思いますが、最近では 2 ギガヘルツ以上のものが出ています。200 倍以上伸びたということになります。

技術革新だけではなく人も増えています。世界の人口はまだまだ増えるという予測があります。増えるのは主に発展途上国だと思いますが、その人たちもネットを使うようになるはずです。How many online という調査ではネット上の人口が 2002 年では 6 億 560 万人ではないかと推測しています。このネット上の人口もどんどん増えていくはずです。ボーンデジタルが増えていく要素はいくらでもあるという状況です。

このような状況で国会図書館ではどのようなことをやっているかといいますと、やはりデジタル情報を大量に作成しています。例えば、近代デジタルライブラリーというものをやっています。これは著作権のなくなった資料や、著作権の権利を持っている方からの許諾を得られた資料をデジタル化してウェブで公開するというプロジェクトです。対象は国会図書館で所蔵している明治期刊行図書ですが、現在では 30,000 タイトル、50,000 冊を公開しています。

収録している画像は白黒の 2 値です。コストなどさまざまな事情により、モノクロのマイクロフィルムを作成し、そのマイクロフィルムをデジタイズしています。このような画像を数百万点収録しています。データ量も相当なはずなのですが、何テラバイトあるのかはわかりません。

WARP というのは、Web Archiving Project の略でインターネット資源選択的蓄積実験事業が正式名称です。こちらは日本国内のウェブ情報を、許諾を得てから収集して保存しようというプロジェクトです。先ほど申し上げました表層ウェブを対象にしています。

2004 年 5 月までに WARP で収集したウェブページの容量は 570 ギガバイトです。たいした量ではないですが、収集頻度をあげると、すぐにとてつもない量になってしまいます。というのは、ウェブページは、普通は更新されるので、更新される度に何度も収集する必要があります。古いウェブページだけあっても仕方がないということです。その更新の頻度の高いものに合わせて収集頻度も上げていかなければならないのですが、高いストレー

ジを使っているので簡単にはストレージ容量を増やせない、そして収集頻度も増やせないという事情もあります。また許諾を得て収集するのでその作業自体もかなりの手間がかかります。もし国立国会図書館が本格的に日本国内のウェブページのアーカイビングをしなければならなくなったりすると、すぐ数十、数百テラバイトになってしまってもおかしくはないと思います。

次に Dnavi についてですが、これはウェブ上に存在するデータベースへのナビゲーションサービスです。現時点では 9,000 件のデータベースのナビゲーションを行っています。ウェブには表層ウェブと深層ウェブがあることを先ほど申し上げました。表層ウェブの収集は、収集ロボットというプログラムを動かして収集することができます。しかし深層ウェブはデータベースです。データベースは収集ロボットで簡単に収集できるものではありません。アクセスするたびに動的にページを生成するので、どのような条件でそのページを表示するかということも重要になってきます。そのためにデータベースへの案内だけはせめてやったほうがいいのではないかという発想で作られたサービスです。条件を入れて検索すると何らかのデータベースが表示されます。

これは電子展示会です。国会図書館が所蔵する資料を中心に、日本の歴史や文化的な資料をデジタル化して特定のテーマで編集し、インターネットのギャラリーとして公開するというプロジェクトです。

これは「日本国憲法の誕生」です。日本国憲法制定に関係する当時の資料をデジタル化して、ホームページで公開しています。 終戦の証書のデジタル化した画像も見られます。

また国会図書館には古い本がたくさんあります。昔の著名人が持っていた本も、いろいろな経緯で入ってきています。その本には蔵書印が押されていたりするので著名人の蔵書印を集め解説を加えて作ったのが、この電子展示会「蔵書印の世界」です。

これは「日本の暦」という電子展示会です。江戸時代に大小暦というものが流行ったそうです。昔は太陰暦だったのでは一月が 30 日の大の月と、一ヶ月が 29 日の小の月が、交互というわけではなく、いろいろ混ざって１年が構成されていたようです。大小暦の中には謎を解かないと、どの月が大で小なのかわからないように作られたものもありました。そのような謎解きが必要な大小暦を集めて作ったのが「日本の暦」というものです。

国立国会図書館は納本図書館であり、保存図書館でもあるので納本された本を保存して後世に伝えていかなくてはなりません。普通の本、雑誌以外にパッケージ系の電子出版物も納本されます。パッケージ系電子出版物とは、国会図書館の言い方なのですが、パッケージ化されたデジタル情報のことです。LD だとか DVD、音楽 CD、パソコンのソフトウェアや電子辞書といったものを指します。このようなデジタル情報も長期的に保存してアクセス可能にし続けなくてはなりません。

デジタル情報は非常に便利です。まず、検索するのが簡単です。作ることも簡単です。配布も簡単です。ファックスを使うよりもはるかに簡単です。ウェブページに掲載するという形の配布形式もあります。使うのも簡単です。パソコンがあれば大体何でも見えて使

えてしまいます。簡単で便利なので、ますますデジタル情報が作られているのだと思います。

デジタル情報をどう保存するのかということも考えなくてはいけません。普通は CD-R、DVD-R などのメディアに保存する、またはハードディスクに入れたままにするのかと思います。しかし電子媒体の寿命は非常に短く、紙であれば保存環境にもよりますが、200 年ぐらいはもつだろうといわれています。マイクロフィルムも相当の長寿を誇っています。よいマイクロフィルムを使ってきちんとした環境で保存すれば、500 年以上はもつだろうといわれているメディアです。それに比べて電子媒体の寿命はどうでしょう。CD-R は 2 年から 30 年もてばいいだろうという話もあります。CD-ROM であってもせいぜい 5 年から 50 年、DVD は 20 年程度、DVD-R も 10 年程度、ハードディスクは 10 年から 20 年、磁気テープは 5 年から 20 年、MO も 20 年から 30 年だろうといわれています。いろいろな考え方によって何年生き延びるかということは違ってきますが、短いということには変わりません。

媒体の寿命が短いだけではありません。媒体の規格も次から次へと出て変わっていきます。フロッピーディスクは 8 インチという大きなものがありました。それが 5 インチ、最近では 3.5 インチとなっていきますが。どうでしょうか、フロッピーディスクは最近あまり使われていないのではないかという印象があります。CD-ROM も似たようなもので R とか RW があります。MO もありますし、DVD もよくわからないほど規格が出ています。DVD の次にはブルーレイディスクが出るのではないかとか、アドバンスド・オプティカル・ディスクといった対抗規格が出るとか出ないかという話も聞いています。このほかにも魅力はあったのに廃れてしまった媒体規格がいくつもありました。

媒体規格がこれほど次から次へと出るということは、規格の寿命も短いということです。デジタル媒体に保存するデジタル情報の寿命が短いのであれば、アナログ保存すればよいと考えられる方もいらっしゃるかもしれません。紙であれば相当持ちますし、マイクロフィルムであればさらに持ちます。しかし、そう簡単にはいきません。マイクロフィルムや紙を利用したアナログ保存が可能なのは印刷可能なものです。動画やインタラクティブなものは紙やマイクロフィルムで保存することは出来ませんし、紙に印刷またはマイクロフィルムに焼いたとしてもあまり意味がありません。ウェブページの保存にも向きません。全てのリンクが切れたウェブページを保存しても無意味だと思います。利用ということを考えますと情報量が増えるほどインデックスのようなものが必要になります。そのようなものは、デジタルで、つまりデータベースのようなものを作ることになりますが、元々デジタルであるのに、検索だけデジタル、使うときはアナログということはあり得ないかと思います。当然、配信用にデジタルの部分を用意することになります。であれば、デジタル情報をアナログ保存するのは、殆んど緊急避難的な場合しかあり得ないのかと思います。

２．デジタル情報の長期保存

デジタル情報の長期保存とはどのようなことなのか。まずデジタル情報が何なのかとい

うことですが、デジタル情報の特徴として次のものをあげることができます。まず、デジタル情報は電子媒体に固定されています。そして再生する機械が必要です。また、媒体には色々な種類があります。中にはOSやアプリケーション・ソフトウェアを必要とするものがあります。このような性質を挙げることができるかと思います。

しかし、デジタル情報を長期保存する場合に、まず考えなくてはならないのは、技術的な依存関係です。非常に雑な整理をしてみます。デジタル情報は媒体に記録されていて、媒体からデジタル情報を読み出すドライブが必要で、ドライブはパソコンに接続され、パソコンにはOSがインストールされないと動作せず、OSの機能を使ってアプリケーション・ソフトウェアが動作します。例えば、PDFの文書がDVD-Rに記録されている場合は、この媒体がDVDのドライブに装着され、ドライブが例えばIBMのパソコンに接続され、パソコンにはWindowsXPがインストールされ、その上でアクロバットリーダーが動いて、ようやくPDF文書を再生することができます。DVDですから当然CD-Rのドライブで読むことはできませんし、PDFなのでワードパッドで開いても意味がありません。WindowsXPなので、例えばMachintosh用のソフトウェアは、普通は動作しません。

この他にも依存関係があります。Windows3.1やWindows95で動いていたアプリケーション・ソフトウェアがWindows2000やWindowsXPで必ずしも正常に動作するとは限りません。動かないものも多数あります。

技術的な依存関係を整理しますと、媒体はドライブに依存して、ドライブはパソコンにつながないといけない、アプリケーションは特定のOSの上で動くといった依存関係です。バージョン間の依存関係というものもあります。この依存関係は技術革新により、ますます複雑になります。

次にデジタル情報を長期保存する場合に問題となるのが寿命です。媒体の寿命、機器の寿命、OSの寿命、アプリケーションの寿命、記憶フォーマットの寿命それぞれを考える必要があります。

まず媒体の寿命ですが、すでに申し上げましたが、せいぜい20～30年だろうと言われています。媒体を再生するためにはドライブが必要ですが、ドライブは機械です。機械なので物理的な寿命があります。ドライブの規格としての寿命もあります。古いドライブを何とか入手したとしても、最新のパソコンにはつなげられないといったことはよくある話です。さらに古い形式のドライブは市場で入手することがますます困難になります。単純に媒体だけとっておいても意味がないということです。

パソコンにも当然寿命があります。電子部品の寿命は意外と短いと聞いております。壊れた部品だけを取り替えたり改造したりして使い続けることもできなくはないことですが、いつまでも可能ではなく、古いパソコンはあっという間に市場から姿を消してしまいます。

OSにも寿命があります。OSは特定の機種で動作します。つまりパソコンの寿命に依存するわけです。古いOSを最新版のOSで動かすということも出来ないことではないのですが非常に大変です。デバイス・ドライバを入手するのが困難だという話はよく聞きます。数

年ごとに新しいOSが登場しそのたびに、古いOSのサポートが打ち切られてしまいます。市場からも姿を消して、いつの間にか古いOSは入手困難になります。オークションなどで入手できることもありますが、メーカーはソフトではなくてそのソフトの使用権を売るという商売を考えているので、他人からソフトのパッケージを譲り受けたからといって、それを使っていいかどうかという判断が必要になります。

アプリケーション・プログラムにも寿命というのもあります。というのは、アプリケーション・プログラムは特定のOSの特定のバージョンでしか動作しないからです。こちらも市場での寿命があります。古いバージョンの入手は困難です。

記録フォーマットにも寿命があります。いろいろな記録フォーマットが次から次へと出てきています。いつの間にか使われなくなったフォーマットというのもあります。ワープロ専用機は市場で流通しなくなったので、ワープロ専用機が使っていた記録フォーマットは、衰退したフォーマットと言えるかと思います。再生のためにはそのフォーマットのデータを呼び出すためのアプリケーション・ソフトウェアが必要になります。記録フォーマットの寿命はそのアプリケーション・ソフトウェアの寿命にもよるということです。

また出た図なのですが、この中のどれかがだめになるだけでデジタル情報の再生ができなくなるということです。

デジタル情報だけ保存していても「何これ？」となってしまうことがあります。それが何なのか、どのような環境で再生するのか、OS の種類、必要なアプリケーション・ソフトウェア等の付随する情報がないと再生ができません。このような情報についての情報をメタデータといいます。適切なメタデータを与えないとデジタル情報を保存しておいても無駄になってしまいます。

デジタル情報は改ざんが非常に簡単なので、改ざんされていないかどうかを示すメタデータも必要だともいわれております。改ざんされたデジタル情報を長期保存しても仕方がないので、改ざんを防止するための技術として電子署名の技術があります。電子署名は認証局を必要とする技術ですが、認証局が長期間存在しないと、そのデジタル情報の正当性が保障されないというのでは、デジタル情報の長期保存のために使うことはできません。認証局なしで改ざんがされていないことを証明できる技術が必要になります。

ある媒体、ある情報だけを小規模に保存すれば良い場合は必要ないのですが、多くの種類、大量のデジタル情報を保存し、利用する機関にとっては、それを入れるための入れ物が必要です。技術的な依存関係や媒体の寿命、メタデータ、改ざんといった課題を何とかする入れ物としてのシステムが必要になります。

全ての機関にはあてはまりませんが、大量のデジタル情報を扱う多くの機関にとって、デジタル情報の入れ物としてシステムの中には、大規模なストレージが必要になります。例えば宇宙観測データのようなものは一ヶ月で一つの観測所で１テラバイト程のデータが作成されるという話を聞いたことがあります。企業によってはペタバイトのデータをすでに持っているところもあるそうです。大規模なストレージが必要だという機関はかなりあ

るわけです。この大規模なストレージのバックアップは非常に大変です。例えば１秒間に100 メガバイト転送できるというのは比較的速い方だと思いますが、そのような転送速度であっても、100 テラバイトのストレージのバックアップやリストアには 10 日以上かかってしまいます。その間システムは使えないので、会社も業務にも相当の支障が出てきます。さらに、ハードの寿命が５年程度でメディアの寿命も 10 年程度ということを考えると、大体 5 年ごとに大規模なデータ移行をしなければならなくなります。そのたびに会社が 10 日間の開店休業状態になるということはあり得ないと思います。

課題を整理してみます。技術的な依存関係です。フォーマットや OS、アプリケーション、バージョン間の依存関係というものもあります。それぞれの技術要素は全て短寿命です。メタデータが無ければ保存していても仕方がない、何だかわからなくなります。データの改ざんについての配慮も必要です。デジタル情報を長期保存するための器としてのアーカイブシステムが必要です。規模が大きい場合はストレージについての配慮も必要になります。

このようなデジタル情報の長期保存という課題に対して、海外では早くから取り組みが始まっています。

まずアメリカです。NDIIPP というプロジェクトが実施されています。NDIIPP というのは、National Digital Information Infrastructure and Preservation Program という非常に長い名前です。約１億ドルという大きな予算がついています。デジタル情報の課題というのはいくつもあるので、それら全てに包括的に取り組もう、さらに全国的に取り組もうというプロジェクトです。協働ということを重視しています。NDIIPP はアメリカの議会図書館が引っ張っているプロジェクトですが、アメリカ国内では様々な機関がデジタル情報の長期保存につながるような研究、取り組みをしています。

オランダも進んでいます。ここも国立図書館が引っ張っています。IBM と共同研究を行い、これをもとに納本システムの e-Depot というものを作っています。

イギリスの場合も英国図書館が引っ張っています。英国図書館だけではなく、DPC（Digital Preservation Coalition。電子情報保存連合。）や JISC（Joint Information Systems Committee。英国情報システム合同委員会。）という機関も様々な研究をしています。

オーストラリアも国立図書館が中心になってデジタル情報の長期保存に取り組んでいます。

では、日本ではどうでしょうか。国立国会図書館は平成 14 年度から調査研究を始めたという程度です。内容は海外文献の調査や、アンケート調査、国立国会図書館が持っている資料の実態調査などです。e-Japan でも関係することを行ってはいるようですが、電子文書の比較的短期の長期保存を考えていて、それを可能とするための e-文書法といったものを作ろうとしているようです。そのほかにもコンテンツ政策の推進ということが謳われていますが、これは国会図書館に政府刊行物のアーカイブを作らせて政府機関のホームペー

ジの収集をさせて長期保存をしようという話です。

再度、課題を整理します。まず、デジタル情報の長期保存を考える場合は、技術的な依存関係に配慮しなくてはなりません。次に、それら技術要素は全て寿命が短いということです。メタデータも必要です。真正性について何かしらの配慮が必要です。入れ物であるアーカイブシステムも必要です。大規模ストレージについても課題があるということです。

駆け足ですが、前半はこれで終わりです。

司会　ありがとうございました。それでは休憩に入ります。

- -

司会　後半に入ります。今野先生、よろしくお願いいたします。

今野　国立国会図書館が実施した調査の一部を簡単に紹介させていただきます。

３．国立国会図書館が実施した調査の概略

平成14年度にアンケート調査を行いました。日本国内の582機関にアンケート票を送り、233の機関から回答をいただきました。回収率は40%です。

「5 年以上保存しているデジタル情報はありますか」という問いに対して「あり」と答えた機関は、232機関中約6割です。「あり」の割合が高いのは、デジタル・コンテンツを作っている機関や大学図書館・新聞社・美術館・博物館といったところでした。

「どのような媒体に保存していますか」という質問に対しては、このような回答をいただいております。「ハードディスクに入れたまま」と、「CD-Rに焼いて保存している」ところがほとんどです。

これは全体にではなく最初のアンケートの問いで「5 年以上保存しているデジタル情報がない」と答えた機関に対しての質問ですが、「これから先デジタル情報を長期保存する予定がありますか、そのような計画がありますか」という問いです。「ある」というのは 4割しかありませんでした。

これは全体に対しての質問ですが、「保存したデジタル情報が使えなくなった経験があるか」という問いです。「ない」が6割少しです。「ある」といった回答のその内訳なのですが、「ソフトがバージョンアップした」、これはよくある話です。「組織内のシステムが変わってしまった」、これもよくある話だと思います。「ハードウェアが古くなった」「保存環境が悪くて媒体が劣化して読めなくなった」、中には「デジタル情報の所在自体がわからなくなって使えなくなった」というところもあるようです。

「公的機関による長期保存の必要性はありますか」という質問です。「公的機関による長

期保存の必要性はある」と回答した機関が約９割です。

アンケート結果を簡単にまとめると、日本国内ではあまり危機感がないということかと思います。公的機関がやることではないか、やってほしいという程度だったと思います。ただ、これは１年半ほど前の状況です。

昨年度は次のような調査もやりました。国立国会図書館が持っているパッケージ系電子出版物が実際に使えるのかという調査です。ただし対象はパソコン用のものに限定しました。対象は 1999 年までに国立国会図書館が受け入れたものです。全部で１万数千点あるのですが予算や期間など様々な制約により 200 点しかサンプルを選べませんでした。これは「無理なことは分かっているが、旧式 OS 用に作成されたソフトウェアを、最新の WindowsXP や MacOSX で再生することができるかどうかという調査」です。結果は、「インストールができない」「インストール途中に失敗する」「アプリケーション・ソフトウェア動作時に異常終了する」「そのデータを読み出すアプリケーション・ソフトウェアがない」ですとか、「媒体を読み取ることができない」、「媒体を読み取るためのドライブがない」ということでした。インストールの失敗やアプリケーションの異常終了は、OS とアプリケーションが適合していないということです。アプリケーション・ソフトウェアは特定の OS で動くことを前提に作られていますので、Windows3.1 用のソフトは WindowsXP では動かないことがあります。アプリケーションが古く、廃れてしまい、入手が困難で使えなかったものも相当あります。「読み取り不可」というのは壊れている、媒体が劣化しているものです。「対応ドライブがない」というのは媒体が技術的に旧式化してしまったからです。最新のパソコンに５インチのフロッピー・ディスクドライブをつなぐのは簡単ではありません。それなりの特殊な技術がないと接続することはできないと思います。

使えなかった理由を分類するとこのような割合でした。OS とアプリケーションとの不適合が原因で使えなかったものは、サンプルの 50%でした。アプリケーション・ソフトウェアが手に入らなかった、古くて入手が困難だったという利用不可理由には、ダウンロードしないと入手できないものも含めています。さらにアプリケーション・ソフトウェアにプラグインをインストールしないと使えないものがあります。そのプラグインのバージョンと使うアプリケーション・ソフトウェアのバージョンが合わないためにアプリケーション・ソフトウェアが動作しなかったものもこの中には含まれています。これが 30%です。媒体というのは媒体劣化が原因、または対応するドライブが手に入りにくいというものです。これが 12%です。その他というのが、再生環境がわからない、メタデータが十分ではなかった、またはセットアップ用の CD が同梱されていなかったというものです。これが 8%でした。

使えない理由を受け入れの年代順に整理したのがこの表です。一番下に 1999 年以前のパソコン用のソフト全体の利用可能と利用不可能の理由の割合を示しています。サンプル 200 点の 7 割弱に何らかの問題があったということです。問題があると言っても、全く使えないということではありません。この調査では使えないと分類したものの中には、必要

なアプリケーション･ソフトウェアをダウンロードすれば利用可能となるものも相当含まれています。また、サンプル数も 200 点と少ないにもかかわらず、パソコン用のソフトは多様です。データだけのものや特殊なアプリケーションが入っているもの、特殊なアプリケーションを必要とするものなどさまざまなものがあります。異なるサンプルが選ばれていれば全然違った結果になる可能性が十分にあるということをご承知おきください。

４．デジタル情報の長期的保存とアクセスのために

後半の本題に入ります。

（１）長期保存戦略

長期保存戦略と呼ばれているものがあります。マイグレーションとエミュレーションというものです。長期保存のためにはマイグレーションとエミュレーションが必要だといわれております。このマイグレーションとエミュレーションの二つで技術的な依存関係とそれら技術の短寿命に対処できると考えられています。

マイグレーションとは何かといいますと、色々な分類ありますが、ここではこのように分類します。電子媒体の寿命が短いことは既に申し上げました。このような寿命の短さを考えると定期的に新しい媒体に移し変えるということが必要になります。つまり定期的な新規媒体への移行が必要だということです。それだけではなくて記録フォーマットも移り変わります。必要なアプリケーション･ソフトウェアが手に入らなくなることもあります。今は Excel、MS-Word 等を普通に使っていますが、将来も確実にそれらを使うことができるとは言えません。そのために記録フォーマットを変換することも考慮しなくてはなりません。記録フォーマットを変換してある時点で一般的なフォーマットに変換できれば、それ以降、暫くの間はデジタル情報の再生が容易になります。寿命が長いと思われる記録フォーマットに変換こと も有効です。ただし、マイグレーションというのはコピー作業でもあり、コピー・プロテクトされた資料のマイグレーションは困難になります。

エミュレーションも長期保存戦略の一つです。エミュレーションというのは再生環境を擬似的に再現しようというものです。古いパソコンのハードウェアやその上で動く OS を新しいパソコンの上で擬似的に再現する、そのためのソフトが売られています。エミュレーションを行うソフトウェアをエミュレータと呼びます。通常は、このエミュレータはアプリケーション･ソフトウェアとして動作します。パソコン上でゲーム専用機の環境を擬似的に再現するエミュレータというものもたくさん作られています。古いパソコン、例えば PC-9800 を擬似的に再現するエミュレータというものも作成されています。消え去ってしまった過去のパソコンのエミュレータもたくさん開発されています。このエミュレーションを使うことでパソコン本体の物理的な寿命を乗り越えようというものです。ただ、エミュレーションすることは技術的に意外に難しいことのようで、エミュレータによっては完

全に動作しないソフトもあります。

（2）OAIS

OAIS についてです。OAIS というのは Open Archival Information System というものです。これはデジタル・アーカイブシステムの参照モデルです。参照モデルというのはある環境における要素間の関係などを理解するための枠組みであって、参考にしてくださいという程度のものですが、これは ISO の規格になっています。非常に大雑把にいいますと、OAIS というのはアーカイブシステムの機能要素とそこで扱う情報の構成を定義している規格です。

これが OAIS の中身、構成です。ピンクの枠の中が OAIS です。OAIS の外には情報の作成者と、情報の利用者、そして OAIS のアーカイブシステムを管理する管理者がいます。このような環境に置かれることを OAIS は想定しています。OAIS は受入、データ管理、アクセス、保存用ストレージ、管理、保存計画といったコンポーネントから構成されます。

OAIS では情報パッケージという形でデータのやり取りと保存を行うことを考えています。作成者からは SIP、Submission Information Package という形で情報パッケージをもらって、OAIS の中では AIP、Archival Information Package という形で保存しておきます。利用者に対してはそれを Dissemination Information Package、DIP といった形で配信するということを考えています。

情報パッケージの構成です。コンテンツ・データ・オブジェクトというものが、もともとの保存対象としてのデジタル情報です。それ以外は全てメタデータです。これがないとだめだろうということを OAIS では考えています。

（3）メタデータ

次はメタデータです。メタデータというのはデータについてのデータです。メタデータといいますと書誌データのようなものを思い浮かべる方の方が多いかと思います。書誌データももちろんメタデータですが、長期保存のためには、保存用のメタデータや管理用のメタデータといったものも必要になります。

コンテンツ・データ・オブジェクトがそもそもの保存対象のデジタル情報です。それ以外は全てメタデータです。情報パッケージはコンテンツ情報と保存記述情報で構成されます。コンテンツ情報の中に、保存対象のデジタル情報がコンテンツ・データ・オブジェクトとしてあり、その他に表現情報があります。この表現情報は OAIS では、コンテンツ・データ・オブジェクトを意味のある概念に対応づけるもの、という少し分かりにくい定義がされています。あまり有効ではないと思います。

この OAIS の情報パッケージに対応するメタデータというものを考えているグループがあります。アメリカの OCLC と RLG という団体が共同で作ったグループなのですが、彼らの考えではこのような OAIS の定義を定義し直して、表現情報というのはコンテンツ・デー

タ・オブジェクトを再生するために必要な OS やアプリケーション･ソフトウェア、CPU の種類、必要なハードディスク容量、必要な周辺機器といった情報を記述する場所としてとらえ直しています。再生に必要なソフト・ハード一式を記述する場所として表現情報を考えているということです。

保存記述情報は長期保存に必要なその他の情報を入れる場所です。参照情報はコンテンツ・データ・オブジェクトを識別するための情報を入れる場所です。普通は何らかの ID です。アーカイブ内における ID かもしれませんし、世界中でのユニークな ID かもしれません。

コンテクスト情報はコンテンツ情報とその環境との関係を記録した情報です。コンテンツ情報が何なのか、コンテンツ情報がヘルプファイルだったり資金調達履歴であったり、何なのかという情報を記述する場所がコンテクスト情報です。

来歴情報というところにはコンテンツ情報の履歴を記録します。コンテンツ情報がどの機関で作られただとか、どこから受け取った、どのように更新したとか、保管者が誰だったのかといった情報です。

固定性情報というところはコンテンツ情報が不正な方法で変更されていないことを保障するための情報を入れる場所という定義がされています。チェックサムですとか電子証明書です。

メタデータの記述要素や構造を定義するものをメタデータスキーマといいます。デジタル情報の長期保存に使えそうなメタデータスキーマが様々な機関で検討されています。METS というのは XML ベースのメタデータスキーマです。先ほどまで何度か説明させていただいた情報パッケージに対応するメタデータスキーマです。情報パッケージを記述するのに METS が使えるのではないかと考えています。その下の MODS というのは、情報の検索に役立つ書誌的な情報を記述するスキーマです。こちらも XML ベースのスキーマです。その下の Dublin Core、これはスキーマとは少し違うかもしれません。メタデータ要素かと思います。その下の MIX というのは画像のスキーマです。

（4）真正性、ストレージ

課題には真正性やストレージもあるといいました。これはよくわかっていません。調査が進んでいないということもありますし、われわれ国会図書館の調査担当者の技術的なスキルの問題でもあります。

例えば、富士通ではオーガニックストレージという、おもしろいストレージシステムを考えているようです。デモを見た限りではバックアップが不要で、規模も拡張可能で、何十台ものストレージ付きの小さいパソコンで構成されるシステムだったと思います。一つのモジュールが故障したとしても常にデータが多重化されていて、さらに分散されて保持されているので問題なようです。故障モジュールがどんどん増えていってもある程度までは大丈夫で、故障したモジュールを取り替えれば問題なく動くと言うものでした。取り替

えられた新しいモジュールには、他のモジュールからデータが分散、再配置され、全体の負荷が均一になるというものでした。意外と使えそうだなあという印象をもちました。

カリフォルニア大学ではオーシャンストアというグローバルストレージの研究がされているようです。これはいろいろ問題があるような取り上げ方をされてしまう peer to peer の技術を使った地球規模のストレージを作ってしまおうという研究プロジェクトのようです。これもきちんと理解したということではなく英語の文献を縦に読んだような理解の仕方なので、本当に使えるのか、今の状況はどうなのかということまでは知りません。ただし、非常におもしろい技術、少し使ってみたい技術ではあります。

真正性については先ほど認証局がないと長期間の真正性といいますか原本性を保障できないと申し上げました。認証局なしで原本性を保障する技術としてヒステリシス署名というものがあるようですが、技術に明るくないので詳しくは分かりません。

（5）対処方法の限界

課題とその対処方法や対策について説明してきました。課題の技術的な依存関係や技術の寿命の短さといったものには、マイグレーションとエミュレーションで対処できるのではないかということ、メタデータについてはいくつものメタデータスキーマが検討されているということ、アーカイブシステムについては OAIS という規格、考え方が使えそうだということ、真正性とストレージについては、よくわかっていない、ご存知であればぜひ教えて欲しいということを申し上げました。

OAIS ですが、これに準拠したシステムがすでに作られております。DIAS というのはオランダ国立図書館とオランダの IBM が共同で研究して作り上げたシステムです。IBM は売ることも考えているようですがよくわかりません。Dspace というものもあります。こちらはヒューレット・パッカードと MIT の図書館が作ったシステムです。オープンソースなのでソース一式をダウンロードして構築するということも可能です。国内ですでに早稲田大学が Dspace を構築したという話も聞いていますし、国立情報学研究所や複数の大学でこの Dspace を使って機関リポジトリの研究をしよう、実証実験をしようという動きもあると聞いています。ただ Dspace というのは、ソース一式は簡単に入手できるのですが、構築することは実は難しいと聞きました。ソースプログラムを読まないとよくわからない、うまく動かせないといった話を聞いています。

OAIS が使えそうだという話をしましたが、OAIS だけでは長期保存を行うことはできません。それは長期保存戦略がないからです。長期保存戦略をどのように適用するかということを考えなくてはなりません。ところが OAIS では長期保存戦略の適用方法については規定していません。規定していないというよりも規定できない性質のものです。なぜならば、それぞれの機関によって扱うデジタル情報は違いますし、考え方も違います。エミュレーション、マイグレーションをどのように適用するか、長期間続けていくかということには、その機関のポリシーがかかわってきます。例えばテキストファイル、テキストデータしか

持っていないような機関であればエミュレーションの必要は全くありません。マイグレーションもハードディスクに入れればすむということであれば、定期的なハードディスクの入れ替えだけで済みます。しかし、さまざまなフォーマットを扱う機関であれば、別の長期保存戦略の適用方法を考えなくてはなりません。

エミュレーションとマイグレーションの適用方法を、各機関ごとに考えなくてはならないのですが、そのためには長期的、全体的な展望が必要です。それは保有しているデジタル情報の種類や量といったものにも関係します。あまり考えずに場当たり的にエミュレーション、マイグレーションを行うこともできなくはないですが、それが長期的な保存にとって不都合なことになりかねません。お金の無駄遣いになるかもしれませんし、や不適切な媒体にマイグレーションしてしまうかもしれません。

オランダの国立図書館が作り上げた e-Depot というシステムではいくつか実験的なことをしています。参考になりそうですがまだまだ決定版ではないという印象です。

エミュレーションやマイグレーションを適用するためには、デジタル情報に十分なメタデータを付与しなければいけません。フォーマットや動作環境についての情報はエミュレーションのためには必要ですし。マイグレーション、例えばフォーマットの変換をする場合では、もともとのフォーマットを知る必要があります。

実際にマイグレーションをやるというのは、実は非常に大変なことです。媒体の寿命が短いので、定期的なマイグレーションが必要です。媒体の寿命やドライブの寿命、OS の寿命を考えて、これらが確実な間に次の環境、次の媒体に移す必要があります。新しい OS の登場頻度や、媒体寿命などを考えると、5 年ごとのマイグレーションが必要かと思います。これも機関のポリシーや扱う情報などに当然影響されます。また、扱うデジタル情報の量は増える一方なので、マイグレーションの規模も回を重ねるに連れて大きくなるはずです。

エミュレーションも実は同じようなものです。エミュレーションをするのはエミュレータというアプリケーション･ソフトウェアですが、アプリケーション･ソフトウェアである以上、特定の OS で動作します。ところが OS はいつまでも使い続けることはできません。それは長期的に動作可能なコンピュータというものがないからです。OS が使い続けられないのであればエミュレータも当然使い続けられません。OS が変わるたびにエミュレータの再作成をするのでは大変です。しかし、JavaVM 上で動作するエミュレータは長期間使い続けられるのかもしれません。

さらに知的財産権への配慮が必要な場合があります。これは図書館などの文化の保存機関に限定される話かもしれません。自前でコンテンツを作り保存する機関にとっては関係ないことかもしれません。マイグレーションやエミュレーションに伴って誰かの知的財産権を侵害する可能性があります。さらに、不正競争防止法が規定する不正競争に引っ掛かるとまではいかなくても、かするようなことをしなくてはマイグレーションができない場合もあります。

メタデータスキーマの検討は進んでいるといいましたが、まだまだという印象です。使えないことはないというレベルだと思います。検討の途中、改良の途中であるだけではなく、メタデータを作るということは非常に労働集約的な作業です。自動的にメタデータを作り出すための検討もされていますが、全てのメタデータの自動生成が可能だということではありません。

デジタル情報の長期保存にはお金がかかるといわれています。まず長期間システムを維持しなくてはなりません。マイグレーションは繰り返さざるを得ず、規模も大きくなり続けます。お金のかかる要素はいくつもあります。お金がかかることはわかっていますが見積もりは難しいといわれています。コスト要素にどのようなものがあるかというコストモデルの研究がいくつかなされていますが、そのコストモデルをもとに実際にコストを見積もるということは難しいのです。

ということで、いろいろな課題があることを申し上げた後に、対応策や解決策を説明しましたが、実はまだまだなのだということを言わなくてはならないわけです。

おわりに

終わりに差しかかって答えがないのでは救いがないので、無理矢理ですが、落ちをつけるために次のような話をさせていただきます。まずは協働、協力が必要ではないかということです。デジタル情報の長期保存にはいろいろな利害関係者が絡んできます。作る人は当然、著作権等を持っています。保存する人もいますし使う人もいます。その人たちの利害調整というのが必要になってきます。いろいろな場面で利害がぶつかります。その利害のぶつかりが長期保存をややこしくしています。うまい仕組みが必要になります。まずは利害関係者の協働がデジタル情報の長期保存に向けて必要ではないかと思います。

デジタル情報の長期保存に関係するような研究をしている研究機関、研究者は実はたくさんあります。そのような研究機関、研究者との協働も重要だと思います。そのような方々がいいアイディアを考えてくれないと長期保存というのは非常にハードルの高いままで終わってしまいます。

保存機関もいろいろあります。図書館や博物館、美術館など、いろいろあります。そのような保存機関同士の協働も重要になってきます。場合によっては共通の規格のようなものでアーカイブを構築できれば、運用コストも安くなるでしょうし構築コストも安くなります。

例えば、先ほどご紹介した NDIIPP では協働をかなり重視しているプロジェクトです。まず全米デジタル戦略諮問委員会というものを作っています。議会図書館の館長だけではなく英国図書館の館長や米国公文書館長、商務省の長官も入っています。チューリング賞というコンピュータのノーベル賞のような賞がありますが、その受賞者も呼んで、二十数人の委員会を作っています。残念ながら日本の国立国会図書館には今のところ、そのような人たちを呼んで委員会を作る余力はありません。1 人の担当者が片手間にしょぼしょぼと

やっているのでたいした事はできません。NDIIPP に戻りますが、全米デジタル戦略諮問委員会の他に、利害関係者会議といったものも作っています。テレビや映画、ラジオ、音楽といった専門協会、図書館、博物館、新聞、雑誌、出版社、あとはウェブデザインだとかシステム開発の企業など、関係しそうな人を包括的に集めて会議を開き、精力的に意見聴取を行い、現状を理解してもらうということを行っています。

NDIIPP では分散保存基盤を考えています。アメリカの議会図書館の中にアーカイブシステムを作って、そこで保存すれば OK だというような発想ではありません。まずデジタル情報の長期保存に協力してくれる協力者を集めてネットワークを作り、それら協力者を結びつける技術基盤を作ることで分散保存を可能にしようと考えているようです。詳しい仕様はまだ公開されていません。

日本の場合はこれからです。まだ何もありません。今日のセミナーが何かのきっかけになればよいと思います。

どうも長い時間有難うございました。

司会　今野先生、どうもありがとうございました。専門的かつさまざまな問題を含んでいることにつきまして、非常にわかりやすくお話いただけたと思います。

- -

司会　それではただ今のご講演につきましてご質問などありましたらお受けしたいと思います。ご質問だけではなくて、今日のテーマにつきまして専門的にご研究されている皆さんもご参加いただいておりますので、ご意見などありましたら自由にご発言いただければと思います。

金澤　金澤です。デジタル情報というのは長期保存するにはいろいろな課題があって問題があって、その答えがないのではないかという気がしていましたので、すごく心配しながら「そうだ、そうだ」と思って話を聞かせていただきました。国会図書館ではパッケージ系電子出版の閲覧を昨年でしたか開始すると聞いていましたが、先ほどの調査の結果、5年以上昔のものについては半分ぐらいが今の XP では見られないというお話がありました。今のような問題が多い中で電子出版物の閲覧対応をするということは、それぞれの OS なりアプリケーションなりを全部準備しないと読めないのではないかという心配をしています。具体的にどのように進めているのかをお聞きします。

今野　国会図書館のパッケージ系電子出版物の閲覧のための体制は、情けないことにようやくこの調査で課題がわかったという状況です。長期保存にはいろいろな課題があるのできちんと対応するには手間もすごくかかります。現時点では今おっしゃったように古いハ

ードやアプリケーション一式をしばらくとっておくしかないのかと思います。そもそもそのような配慮が必要だということをきちんと認識していなかったということです。

国会図書館はそれなりに大きい組織でわたしは関西にいて東京の状況をよくわからないまま委託業者に調査を依頼したということです。東京の資料所管部署の担当者と具体的な打ち合わせを何度もしているのですが、東京に出張する時間や予算がないといった事情がありまして、実際どのような閲覧体制なのかということまでは具体的に私自身が申し上げられるほど理解はしていない状況です。

このような調査結果が出たわけですから近々に何らかの対処をしなくてはならないのですが。やはり長期的な保存ということを考えなければならないという基本方針があって、では短期的にどうするのだ、中期的にどうするのだといった組み立てが必要になるかと思います。そのためのガイドラインを今年は作ろうということで動いております。

司会　さまざまな課題が明らかになって中長期のガイドラインをこれから策定されるということのようですね。

小川　国際資料研究所の小川と申します。私、昨年の暮れに『電子記録のアーカイビング』という本を出したのですが、本当はあまりよくわかっていないのでこれで大丈夫なのでしょうかという意味で上梓させていただいたような次第です。

媒体の寿命、機器、パソコンなどの寿命、OS の寿命に関連して、ドライブと機器の違いを教えていただけますか。もう一つ情報パッケージについて。情報パッケージの図があって、例えばコンテンツ・データ・オブジェクトと書いてありますけれども。これは本だったらページの中に書いてある内容のことをいっているのではないかと思ったのですが。そうするとその表現情報は何になるのかとか、右側の保存記述情報というところで ID が入る、これは図書館だったら請求番号に当たるのだろうかというようなことを考えながら聞いたのですが、そのような対応関係でお話ししていただけたら、ありがたいのですが。

今野　表現がまずかったかと思います。ドライブとパソコンと分けて説明して資料を作れば良かったのかと思いました。そのくらいではだめですか。

二つ目ですが、情報パッケージ、コンテンツ・データ・オブジェクトが何なのかというご質問で、具体的にデジタル情報でないものにどう対応付けられるのかということだったかと思いますが。対応付けられないと思います。全く別物と考え直していただいたほうが、新しいものとして考えていただいたほうがすっきりするのかと思います。仮に本をデジタル情報と対応付けた場合は、本そのものがコンテンツ・データ・オブジェクトになるかと思います。本の場合それを表現するための情報は特に必要ないわけですね。本があればそれで足りてしまいます。保存記述情報に対応する参照情報は確かに図書館の請求記号と対応付けられないこともないなというレベルです。これはやはり情報パッケージの特定や個

別化といいますかユニーク性を保つための情報です、参照情報というものは。別物と考えていただくほうが良いかと思います。

小川　別物と考えられるといいのですが、実はいつも物理的に目に見えるものを考えながらイメージアップするものですから、コンテンツ・データ・オブジェクトが本そのものだとした場合は、保存記述情報にかかわる右側の欄というのは外側から、あとで本を見ながら付けていかなければいけない情報と考えればいいのでしょうか。

今野　それは間違いないです。

小川　そうするとそのコンテクスト情報というのは、例えばシリーズになっているような本の場合、これは何のシリーズの何番目とか、雑誌だったら何年何月号に当たるような、連続性についての情報をコンテクスト情報というように見ればいいのでしょうか。

今野　いま例示していただいたものは全て書誌的な情報の一種だと思います。そのようなものは情報パッケージの外に置くというのがそもそもの考え方のようです。情報パッケージの外に、そのような記述情報、記述メタデータを用意して、コンテンツ情報というか情報パッケージを特定できるのではないかという発想です。コンテクスト情報にはそのような書誌情報相当のものは入れることはあまり考えていないようです。まずコンテンツ・データ・オブジェクトがどのようなものなのか、それを説明するための情報を書くところです。

小川　そうするとコンテクスト情報というのは何が入っているのですか。

今野　デジタル情報は０と１ですよね。それがまず何なのか、これは何の記録なのかという説明が必要なわけです。それは例えば気象観測データで数字が一杯あるのか、どこかの星、火星に探査船を送ってそこから送ってきたデータなのか。そのような説明がないとそもそもわからないわけです。再生するための情報を表現情報に入れるといいのではないかというアイディアはあるのですが。まず人間が理解する人間向けの情報ですよね。これはそもそも何なのだ、画像なのか音声なのか、といったものを入れる場所だと理解しております。

司会　小川さん、よろしいですか。

小川　まだ全然わかっていません。すみません。
保存記述情報の部分は人間がわかるための情報が入るということなのですか。

今野　それだけではないです。例えば固定性情報なのですが、これは改ざんされていないことを示す情報で、これは普通人間が理解できない情報です。何かしらの書き方というかエンコーディング方法があって、人間が見られないことはないですが、見てもよくわからないアルファベットなどの文字列だったりします。来歴情報は、これは比較的人間向けの情報かと思います。どこの機関が作って、いつ受け入れて、何回マイグレーションしたとか、今の保存機関はどこだ、過去の保存機関はどこだ、そのような情報があるほうが長期保存には有効だろうということで作られた情報だと思います。コンテクスト情報、これもそうです、人間向けです。参照情報というのは、これは ID 相当の URI 等いろいろあるようですが、そのような情報を入れておくところです。これは人間が見ても普通は仕方がない情報です。人間向けとか何とかということではなく、そのようなものなのだと理解していただくほうが良いのかと思うのですが、どうでしょう。

小川　おぼろにわかってきたことは、そもそも、この保存されるべき電子情報というのは目に見えないのだという前提で今お話をされています。ではそれが目に見えてきたらということではなくて、目に見えないという状態でどのようなものが入るかということをご説明いただいたということがわかってきました。そうすると目に見えやいけれどもこのようなものが入っていますというのが、コンテンツ・データ・オブジェクトそのものです。それに対してそれを再生するのに必要な情報というのは表現情報に入るわけですね。これが音楽だとか、これは絵だとか。

今野　それがコンテクスト情報です。

小川　それでは OS、OS アプリケーション、機械。

今野　表現情報を再生するために必要な条件を記述したらいいのではないかという考えが最近出ているというところです。

小川　ということは少し前の技術的な依存関係を示した層の図を見せていただいたのが 4 ページにあったのですが、この辺の情報が入ってくるということなのでしょうか。

今野　そうです、そう考えていただいたほうが良いかと思います。

小川　だんだん見えてきました。ありがとうございました。

司会　では、山本先生。

山本　国立情報学研究所の山本と申します。大変立派なお話でした。実はこの問題に昔から興味がありまして。3 年ぐらい前に国立情報学研究所に移ってきましたが、最初にいったことが国立情報学研究所も国会図書館に続いてデジタル・アーカイブをやらなければいけないのではないかという提言でした。最近だんだんそのような話にもなっているのですが。日本のお役所の会計全体が、日の丸に一つあればいいという考え方なのです。要するに一つのところに全部任せる、ほかのところはまたそのほかのことをやると。協働ということをご紹介いただきましたけれども、これが非常に重要なのはデジタルの情報というのはそうでなくても集中しがちです。ところが集中してしまってそれがなくなってしまったらどうしようもないということなのです。図書の場合はたくさんばらまかれますからどこかに残っている可能性があります。あれだけ戦乱が続いてもやはり残っている本のほうが多いですね。それに対してデジタルのものというのは、例えば会社がみんな集めてしまっていて、その会社が潰れたらどうなるでしょう。デジタルの情報を誰も信用しないのはあたりまえです。だから信用させるためにはきちんとしたところがきちんとアーカイブする、しかもそれが一つではなくていろいろなところでやっているということが非常に重要だと思うのです。

今日のお話に本当に感銘いたします。実は今学期、デジタル・ドキュメントという総研大のドクターコースの授業でデジタル・アーカイブの話をしました。国会図書館のサイトも勉強させてもらいながらやったのですけれども。大体同じ範囲を話したにも関わらず、このお話にはとても及ばない話で、わたしの代わりに授業をしていただけば良かったと思いました。どうもありがとうございました。

瀬岡　富士写真フィルムの瀬岡です。わたしもデジタルに関しては非常に興味を持っていますし、今日のお話は現時点でデジタルはどこへ行こうとしているのかというのがわからないということが非常にわかりました。

アンケートのところで非常に興味のあるデータが出てきたのですが。過去のデータを引っ張り出すのに無理矢理今の OS でやるというところがありましたね。それで出せなかったものの順序からいうと OS アプリケーションがあって。媒体も入っていた図があったと思うのですけれども、1992 年だけ特別、緑ですよね、50%ぐらい出せないという。これは偶然なのですか、それとも何か関係あるのでしょうか。

今野　細かい調査結果を見ればわかると思うのですが、手元に資料がないので詳しいことは申し上げられません。まずサンプル数が非常に少ないので、たまたまということは十分あるかと思います。10 年で 200 点、毎年たった 20 点ぐらいのサンプルなわけです。選ばれたサンプルがたまたまそうだったという可能性も十分あります。これは媒体ということなので劣化しているという場合もありますし、5 インチＦＤというのがかなり対象になっ

ていたと思います。そのような理由でたまたまサンプルがそうだったからだったのではないかと思います。もっときちんとした調査を、もっと包括的に全資料を対象に調査ができれば、ここに表示しました割合についてもう少し参考にしていい情報が出せたのかもしれませんが。サンプル数の少なさで極端な差が出た可能性が高いのではないかと思います。

本多　東京大学情報基盤センターの本多と申します。わたしも仕事柄メタデータだとか何だとかという話には結構かかわっていますけれども。実はOAISのことはあまり知らなくてそのような団体もあったというぐらいの認識しかなかったのです。デジタル・コンテンツ、デジタル情報の保存のためにまたメタデータをつけるという、そのような観点はわたしは普段あまりなくて、なるほどと思ったのですけれども。ただこのメタデータ自体がデジタルデータではないかと思うのです。

人間が見てわかる、わからないというのは、単にファイルを開いたときにXMLであれ何であれ、ほぼベターなテキストのファイルとして見えるか見えないかという話ではないかと思うのですが。そうすると結局はメタデータ自体が重要なデジタル情報で、それがなければもとのコンテンツは本当に何これということになるということだと思うのです。実際にはメタデータもどのような形式にせよ現実的にはデータベース化して持っているということになるのだろうと思うのですが。そうするとOAISでは、デジタル・コンテンツの長期保存のためのメタデータはこのようなことをつけていきましょうということを決めていたとしても、OAISのホームページを見ればわかるのかもしれませんが、ただそれに対するメタデータの保存というところではどのようなことになるのでしょうか。結局はメタデータの分散保存ということになるのか、あるいは実際にはメタメタデータというような言葉もありますけれども、そうなっていくのでしょうか。保存という観点からメタデータを考えコンテンツを保存するためにというところで考えたときに、逆にそうするとそのメタデータ自体はどうなのと気になるのですが。

今野　非常に嬉しい突っ込みだったと思います。ご質問の内容を少し変えてしまうことになるかもしれませんが、メタデータの保存ということとメタデータの可読性にかかわるご質問だったのかと。長期保存するということであればメタデータももともとの保存対象のデジタル情報も同じです。一緒に情報パッケージとして保存してしまおうというのがOAISの考え方です。保存したとしても見えなくなってしまうのではないか、形式がよく変わるという話をしたあとなので、メタデータの長期保存、保存したことで見えなくなってしまうのではないかというご指摘と考えたほうがいいのかなと思いました。まさにそのとおりなわけです。

どのような保存がいいのかといいますと、テキストファイルのような形であればアプリケーションは何でもいいと、読めばわかると。実はそのテキストファイル以上にXMLのほうが強いフォーマットだという話を聞いたことがあります。XMLだと文字コードまで指定

して記述できるという説明を聞いた記憶があります。文字コードが違ってしまうとテキスト形式であると今度は読めないわけですね。XML で書いた場合はその辺まで対応できるということだったと思います。メタデータも XML で書くといいのではないかと私は思っています。長期保存に使えそうなメタデータスキーマの検討が進んでいるといいましたが、XML で考えているところがいくつもあります。

METS というメタデータスキーマは XML のスキーマです。XML で情報パッケージというか、メタデータを全部書いてしまえば少なくとも人は読めるわけです。スキーマの記述規則ですとか理解の仕方のようなものをとっておけば、それを人が読むことはできるわけです。それと一緒にコンテンツ・データ・オブジェクトも何らかの形で関連づけてとっておけばいい。それができるのが METS というものであるわけです。

ということでメタデータの保存というかその可読性については長期的にはどうなるか確かにわからないです、XML であっても。少なくとも中期的には十分いいソリューションなのかと思います。

司会　では、そろそろ時間ですので最後のお一方。

佐藤　佼成出版社の佐藤と申します。スキーマについて教えていただきたいのですが。Dublin Core の場合は記述の要素のセットとして決まりごととしてあるだけで、その中でどのような形で記述するかということについての決まりというのはなかったですよね。それが情報パッケージに関して、どのようなスキーマを使うにしても、ボキャブラリーのレベルとか書き方の文法のレベルというのか、そのようなところで決まりごとというのはあるのでしょうか。最終的にはそのようなところで情報を引き出せるのか引き出せないかというのが係ってくるのではないかと思うのですが。

今野　Dublin Core で書くというのはデジタル情報を探すために嬉しいようなタイトルですとか、誰が作ったとか、いつ作られたとか、そのような情報を書けるメタデータの要素だったと思うのですが。これは情報パッケージの外です。ただ METS の考え方ではそれも情報パッケージの中に入れるということができるのですが。

佐藤　それはボキャブラリーのレベルでの決まりごとというのはあるのですか。

今野　METS ですか。私の理解が不足していて。もう少しかみ砕いて質問していただけると。

佐藤　一つ一つの記述の要素がありますよね。そこで使われる言葉の決まりごと。

今野　METS の場合ですか。XML のスキーマですので当然属性として入れる値が何だとか、

値としてというか内容として入れる､値というより何を入れるかというようなルールです。そのようなものはスキーマ定義のほかにも書き方などで規定はされております。

佐藤　ありがとうございました。

司会　それでは議論もつきませんが、ここで第２回のＪＨＫオープンセミナーを終了させていただきたいと思います。改めまして、今野先生、どうも長時間お疲れ様でした。有難うございました。

　私どもＪＨＫでは、今年末ごろに第３回のオープンセミナーを計画しております。今後とも私ども情報保存研究会をよろしくお願いいたします。今日は長時間どうもありがとうございました。