# 電子媒体による公文書等の管理について

杉本　重雄
筑波大学　図書館情報メディア研究科

**背景**
- **私たちは将来のコミュニティのために、紙媒体、電子媒体の違いにかかわらず、文書と記録を残していかねばならない。**
- **インターネットや Web そして電子政府に代表されるように、私たちの情報環境はこの10年余りの間に劇的に変化した。**
- **情報環境が変化したとしても、私たちは将来に向けて記録を残していかねばならない。**
- **一方、電子的な文書・記録を残していくには、数多くの問題を解決していかねばならない。**

## １．背景

最初に背景です。紙媒体とか電子媒体の違いにかかわらず記録を残していかねばなりません。現在、毎日いろいろな情報をつくり出しますし、いろいろな記録をつくり出します。これをいかにして残していくかというのが、将来への我々の責任です。電子政府とか、ウェブ、あるいはインターネット、そうしたことで代表される情報環境は、この15年ほどの間で劇的に変化しました。

大学におりますと、常に接する相手は同じ年代です。大学の１年生、２年生ぐらいで20歳までの学生です。彼らが生まれたのは、今からちょうど20年か19年ぐらい前。そうすると、今の１年生だと全く平成の時代です。まだポケベルの時代だったかもしれません。携帯電話よりは少し前、物心ついたときには携帯電話やインターネットを使うのは普通の世代になっています。彼らに、ネットワークのない生活というのを想像してごらんと言っても、大体が想像できない。皆さんの年齢は様々ですが、15年ぐらい前の記憶は定かな方々ばかりかなと思いますその頃と今を比べてみて、本当に変わっているということは十分おわかりと思います。この変わってしまった情報環境の中で作り出しているいろいろな記録を残していかなければなりません。それが我々の背景です。

本というのを考えてみましょう。ここでは辞書を例にとっていますけれども、冊子体の辞書、それから携帯型の電子辞書、それからネット上の辞書とあります。これは、英和辞典、あるいは国語辞典でも結構ですし、あるいは百科事典でも結構です。こういう辞書、あるいは事典と言うと、冊子体を思い浮べるのか、携帯

杉本 重雄（すぎもと しげお）
筑波大学図書館情報メディア研究科教授・同大学知的コミュニティ基盤研究センター長。現在の研究領域はディジタルライブラリ、メタデータ。Dublin Core Metadata Initiative の評議委員会（Board of Trustees)、ディジタルライブラリに関する国際会議等国際的な活動に積極的に参画。

**本を考える**
**・本って何？**
**・あなたにとって辞書はどれ？**
**・本とは、「もの」それとも「中味」？**
**・両方が一体化したもの（たぶん）**
**ーものとしては、使い勝手や見易さを決める**
**ー中味の情報は、別に紙に印刷したものでなくてもかまわない**
**・ディジタルになると何が変わる？**
**ー「もの」と「中味（コンテンツ）」が分けられる**
**→　同じ中味をいろんなメディアに入れられる**
**ーいろんなものを一緒に入れられる（テキスト、ビデオ、オーディオ、リンク）**

型の辞典を思い浮かべるのか、はたまたネット上の辞典を思い浮かべるのかが問題なんです。

辞書というのは、ある意味典型的な本だろうと思います。確かに小説とはかなり違いますが、我々が情報を得るためによく使うものです。ところが、ある意味で、辞書のイメージがこんなに変わっているんです。ふだん、確かに本と言われると冊子体のイメージかもしれないんですが、現実に使っているものは携帯の辞書であったり、あるいはウィキペディアであったりします。携帯からアクセスすることもあります。あるいは携帯そのものに辞書が載っていたりします。

ですから、こういう環境の中に生きていて、本というのは何なんだろうと考えてみたいです。今、多くのみなさんが辞書に対して冊子体のイメージで持たれているということは、それはやはり本をモノとしてとらえていることになると思います。でも、実際に使うほうからみると、モノというよりは中身で利用されていますね。それが実際のところかなと思います。

本とは何ぞやというのは、自分自身では、基本的にはモノと中身の両方が一体化したものなんだろうと思います。実際には、モノとしての使い勝手だとか見やすさ、アクセスのしやすさですとか、手軽さとか、内容によって本に対する見方が違うと思いますが、でも紙でなければならないということは全くなくなっていると思います。

この後、お話しします電子化の観点からは、ディジタルになって何が変わるかという問題があります。例えば、ディジタルの文書といっても、コンパクトディスクや DVD に入っているものもあれば、ネット上にあるものもあります。ただ、基本的にディジタルになると何が変わったかというと、モノと中身、少し違う言い方をすると、入れ物と中身が分離されたということだと思います。

印刷物というか、冊子体の場合は、インクが紙ににじみ込ませてあるわけですね。ですから、入れ物である紙とインクであらわされた情報というのは分けることができません。しかし、ディジタルデータの場合は、当たり前ですが、いろんなところにコピーが作れます。コピーのしやすさ故に、コピーライトに関連するいろいろな話題がでてきます。でも、基本的にはディジタル化とは入れ物と中身の分離であると思います。

加えて、ディジタルであれば、テキストであろうが、ビデオであろうが、オーディオであろうが、リンクであろうが、いろんなものを同一の基盤の上にあらわすことができる。それは、紙のも

のではできなかったことですね。

ですから、我々人間が使う情報表現というのは文字だけではもちろんありません。いろんなものを使います。ですから、人間が行ういろんな情報表現を、１つの基盤上に乗っけられるというのがディジタル化であると思います。そういうふうに考えると、いわゆるディジタル情報技術は、コンピューターやネットワークを使って、いろいろなものを表現し、伝えるという活動に対して、大きな自由度を与えてくれているのかなと思います。

**コンテンツ（content（s））ということば**
**・コンテンツを、ディジタル形式で表現し、配信し、利用する**
**－ネットワーク（あるいは他のメディア）を介した配信**
**－利用者とその環境に応じた利用を可能にする**
**・ディジタル化は入れ物と中身の分離**
**－中身だけで流通させることができるようになった**
**－ただし、利用のためにはそれなりの道具・環境が必要**

## コンテンツ（content(s)）ということば

コンテンツという言葉は、随分以前から使われている言葉です。やや硬いかもしれませんが、コンテンツという言葉は、訳してしまうと単に「中身」になります。ある種、ふだんよく使うのだけれども、意味がわからない言葉かなと思います。そこで、辞書を幾つか見てみました。基本的には何らかの情報表現をした内容であるというか、あるいは情報表現されたものであると、自分自身は考えています。

単にどこかコンピューターのディスクの中に乗っかっている、ただそれだけでコンテンツと呼ぶのは余りよくないと思います。何らかの意図のもとに表現したもの、それがコンテンツだと思います。ディジタル化することによって入れ物と中身を分離できますので、それで、何らかの入れ物を使って表現されたものの、その中身の部分というのをコンテンツと考えています。

ただ、もちろんコンテンツというものは、そのものでは、見ることも感じることもできませんので、利用のためにはそれなりの道具、それから環境が必要になります。それが従来の出版メディアというんでしょうか、あるいは表現メディアと少し違うところかなと思います。

**情報環境の進化**
**・文書や記録を作り、利用し、蓄積するための環境は劇的に進化し、さらに変化を続けている。**
**－メディアの変化**
**・記録メディア**
**フロッピィ（８、5.25、3.5インチ）、光磁気ディスク、CD、DVD、VHS 等**
**固定ディスクの容量－メガバイトからテラバイトへ**

## 情報環境の進化

情報環境の進化として、はじめにメディアの変化です。皆さん、８インチのフロッピーというのは覚えておられますか。もう全く見ないですね。今の学生は、５インチのフロッピーも全く見たことないです。そろそろ3.5インチのフロッピーも危ないです。USBメモリーもどうなるかわかりません。そういう意味では、フロッピーで８インチ、５インチ、3.5インチ、光磁気ディスク、CD、DVD、VHS、この中で残っていくものもあれば、消え去っていくもの、全く消え去ってしまったものもあります。

自分の学位論文は８インチのフロッピーに入っています。自分

にとっては幸せなこと（？）に、そのフロッピーの中身は決して読めません。１つの理由には、ハードウエアがないこと。中身は非常に単純なので、中身を取り出せればテキストだけです。読みだすことさえできれば何とかなると思います。ファイルの管理システムも CPU も、どこかのコンピューターの博物館に行けば生きているかもしれないです。けれども、身の回りにはないですね。

こうした古いもの、読めなくなっているものは幾らでもあると思います。もちろんあるコンピューターで扱うファイルだけではありません。ビデオテープなんかもそうですね。例えば20年前とか、もっと前に撮ったビデオがあるんだけれども、劣化して中身が読めないとか、８ミリのビデオテープのようになくなってしまったのもありますね。いろんなメディアの変遷の影響を受けますね。

ディスクの容量という問題もあります。これは、保存という話とはちょっと違うのですが、さっき言った８インチのフロッピーというのは、大体256キロバイトぐらいだったんです。今、キロバイトという単位は聞かないです。メガの単位でも、聞かなくなっていました。

3.5インチのフロッピーが1.4メガあり、十分入るなと思っていたんですけれども、今ではメガバイトは、メモリー大きさの単位ではなくなってきましたね。この20年の間にそれほど我々の環境は変わっていっています。

**ーオーサリングツールの変化**
- **タイプライターからワープロそして DTP へ**
- **マルチメディア、ハイパーメディア資料のオーサリングツール**

文書をつくるオーサリングツールにしても、タイプライターからワープロ、そして DTP に変わっていきました。DTP という言葉も、もうほとんど使わなくなったように感じます。デスクトップパブリッシングの略です。デスクトップパブリッシングが普通になってしまいました。

1980年代の終わりごろにマッキントッシュのハイパーカードというソフトが一世を風靡しました。それ自身はもうみなくなりました。そのハイパーカードでつくられていた文書は、カードをベースにして、その中身と中身をリンクで結んでいくものです。今我々が扱っているウェブでは全く普通のものになっています。テキスト同士だけを結ぶのではなくて、絵と絵を結んだり、テキストと音を結んだりと、いろんな形で結べるようになってきました。現在ではそうしたリンクを普通に使っているところが大事なところかと思います。そのころは、いわば特殊なものであったのですが、今は誰もが普通に使っています。普通に使っているということは、我々がふだんつくる文書でも、それを何の気なしに使ってしまっているということかと思います。

例えばマイクロソフトの Word で文書を作っているとき、文

書の中で URL を入れて、Enter キーを押すと、テキストの色が自動的に変わってしまい、しかもアンダーラインがつきます。これは何だろうと思っていたらリンクになっていたといった経験はおありであろうと思います。

こういうリンク付きの文書を保存していこうと考えたとき、それがリンクという情報を、URL という文字の内容で保存しておけばいいのか、あるいはそのリンク先がどうなっているかということの情報までちゃんと保存しておかねばならないのか、という問題があります。この正解はわからないですね。多くを求めれば切りがないです。例えばその URL の文字列だけがあって、それが何ぞやというのは、例えば30年後の人、あるいは50年後の人にわからないでしょう。ですから、我々がふだん使っているソフトが、昔では考えられなかった内容を、普通に、気がつかない間につくりだしてしまうことがあります。

**ー通信基盤の変化**
- **高速ブロードバンドネットワーク**
- **安価なパソコン**
- **携帯電話、携帯端末**
- **Web と電子メール**

次に、通信基盤の変化ですね、高速のブロードバンドネットワーク、安価なパソコン、これはもう我々にとって普通のものになってしまいました。最近ですとネットブックと呼ばれるマシンが数万円の値段です。携帯電話もありますね。

通信系の古い機材の話なのですが、音響カプラというのは皆さんご存じですか。かなり上の世代になると覚えておられるかと思います。要するに電話、アナログ電話を使ったデータ通信のための機材です。

例えば、情報検索するというと、そういう機材を使って、データベースを運営しているところに電話をかけて、それでコンピューターにつないで検索をしていました。コンピューターの使用料も高いですいし、それから回線料も高いですから、できるだけ検索効率をよくしましょうということが非常に大事だったわけです。ところが今、検索というと「グーグル一発」とか、あるいは携帯からでもできてしまいます。これもやはり情報環境の大きな変化です。ですから、そういう環境の中で生きているのであるということをやはり前提に考えておかないといけないと思います。

**・90年代におけるインターネットと Web の爆発的拡大**
**ー情報資源を発信し、アクセスするための主要メディアとしてのインターネット**
**ー大量かつ多様な情報資源を提供し、利用者の情報アクセスを支援する重要なサービスとしてのディジタルライブラリ**

90年代はインターネットが爆発的に広がりました。日本国内では、例えば90年代からは日本政府は e-Japan ですね。2000年代に入って u-Japan に変わりました。ｅからｕ、ｕはユビキタスです。最近、アップルの iPhone のような、いわゆるスマートフォンが出ています。日本や韓国というところでは、携帯電話を使ってネットにアクセスする、あるいは携帯電話の上で発信というか出版をしていくような動きが非常に進んでいます。

例えば、図書館での事例ですが、携帯に対して OPAC（オン

ライン目録）という目録検索サービスを携帯電話向けに提供するということがごく普通に行われています。これは、アメリカへ行くと必ずしもそうではありません。聞き伝えで正確ではありませんが、ヨーロッパだと行われているそうです。

携帯電話をネットワークにアクセスする道具として普通に使ってきているわけです。例えば携帯、もう７、８年前ですね、６、７年前かな、大学図書館に勤められている図書館員の方で、その人から携帯電話でサービスするといいんだよ、と。その理由がおもしろかったからおぼえているんですけれども、端末の数が減らせます。皆そこで自分で持っている携帯で探してくれて、それをそのまま書架の間に入れてくれる。だから、メモ用紙と一緒なんです。

そういう話を聞いて、僕はこれはなるほどなと思って時々紹介している話なんですけれども、多分、ふだん携帯を持っていて、自分に、いつも僕は余り携帯使わないんですけれども、実は一緒にいます。ですから、そこにいろんなものが残せます。ですからそこを、要するに情報を探そうとしている人には、今そこにいるところで情報を得られることというのは、やはり便利なんです。ですから、そういう道具を用意してあげるということは大事なことだし、そういうふうにやっぱり世の中全体に動いていっていると思います。

**・電子政府（e-Government）の潮流**
**－国、自治体**
**－ワークフローやビジネスプロセスの変化**
**－政府行政情報資源への主要アクセスポイントとしての Web の利用**

それで、電子政府の潮流ですね。これも今ほかで、特にこちら国立公文書館、あるいは内閣府では公文書管理法というのができ上がって、それで90年代に e-Japan が進められました。いわゆるｅガバメントですね。その電子政府では、特に文書の流通が電子化していくことになります。書類づくりに付随するいろいろなコミュニケーションも、どんどん電子化されていると思います。

もう一点、例えばこの１年以内に、あるいは２年以内に、オフィスで使う書類をワープロ、あるいはエクセルといったソフトを全く使わずにつくったことはあるでしょうか。ふだんの文書づくり、あるいは文書のやりとりの中で、そういう電子的なものがベースになっていないことはないと思います。そうすると、例えば、決裁を受けた正規の文書としては紙のものであるのかもしれないのですが、文書として最初につくられて、かつそれで実際に流れていくのは、もうほとんど電子的であるとすれば、その電子的なものがメインになって、紙の文書が、ある種補助的なものになっているとしても別に何の不思議もないですね。

それと、私自身は紙でもらった書類を後で電子的に送ってくれませんかとよく頼みます。これは読むためには紙を使うけれども、

・ペーパーレス環境
ー電子的情報資源が主要な情報源となった
・読むためだけにメールや Web ページを印刷し、保存は電子的に行っている

読んだら紙のものは捨ててしまって、とっておくのは、自分が使っている PC の上に乗っかっているものになっている。こうしたことは、結構共通しているのではないでしょうか。ある意味では、紙の消費量というのは減らなかったけれども、仕事の仕方はペーパーレスの方向に向いていると思います。だから、保存は電子で、紙のものは読むだけ、それが普通の仕事の中でかなりの部分を占めていないかなというふうに思います。

それから、加えて、もちろんハイパーリンクですとか、あるいは動画ですとか、音ですとか、そうしたものというのは、紙だけでは表現できませんので、そうした情報も含めて一つの文書と考えると、それはもう全く電子とかディジタルじゃないと扱えないですね。でも、ふだん自分たちで、自分自身がつくったり、あるいは接しているもので、本当に紙とインクというと、今では紙とトナーですかね、それだけであらわせるものというのはどの程度なんだろうなと考えてみると、自分自身で評価したことはないんですが、音ですとか、あるいはリンクというのに頼っていると思います。

ー電子的な情報資源は必ずしも印刷できない
・現代のディジタルドキュメントは、すでに従来の印刷指向の文書とは同じではなくなっている
・統一的な環境で、ビデオやアニメといった要素までをも含むマルチメディア、ハイパーメディアの情報資源を利用することができるようになった

最近、CD の売り上げがずっと下がっているというのがニュースで出ていましたが、それに対して、いわゆるネット上での曲、音楽、楽曲の販売がどの程度になっているのか、そこのところまでちゃんと知らないのですが、ただ、自分自身 CD を買っても、大体、例えば mp 3 でプレーヤーやパソコン上で聞いています。ですから、以前は CD から直接再生をしていたんだけれども、今は、コピーをとってから聞いていないかなと思うんです。ですから、CD は内容を直接楽しむためとか、内容を使うためのメディアではなくて、CD は内容を運ぶための入れ物に変わっているわけです。それが進んでいくとネット上での発信です。今は音楽から進んでいますね。日本だと着メロから始まった着うたですね。世界的にもネット配信が進んできています。便利がよくて安ければ、大体そちらのほうに向いていくと思います。例えば、まだこれからのことかと思いますが、漫画のネット発信のように、出版にしても同じようなことが考えられます。もし紙で印刷して出版するコストが高く、十分ペイしなければ、そちら側が消えていきますでしょうし、ディジタルなものがそれなりに人気が出れば、そちらに向いていくことになると思います。

Google Book Search ですとか、Amazon の Kindle の影響があります。日本国内での経験では、いわゆる電子ブックリーダーというのは何回か出てきては失敗して、うまくいかなかったのですが、なぜか Kindle はアメリカでそれなりに広がりました。基

本的な理由は、電子ブックリーダーとネットワークがつながって、コンテンツをどこででもダウンロードできるというのが、大きな違いかなと思います。ですから、内容を運搬するためのメディアを使わなくても、どこででもアクセスできるというところが大事なポイントだったのかなと思います。そういう意味では、我々の携帯電話の利用と同じことかもしれません。

一方で、漫画の話なのですけれども、少年向けの週刊誌のビジネスモデルに寿命が来ているというお話を聞いたことがあります。それは雑誌そのもので赤字が出てもコミックスでカバーしていたのに、コミックスの売り上げも落ちてきている。そうするとこのビジネスモデルは立ちいかなくなるという話を聞いたことがあります。そのため、次のビジネスのことを考えなければならないそうなんです。そこでディジタルな漫画ということになるのだけれども、そのマーケットは伸びてはいるが、まだ小さくてというお話しでした。

ディジタル化した本から冊子体の本を作ることも、もちろん可能です。ミシガン大学の図書館で実際に見せてもらったことがあるのですが、ディジタル化したコンテンツを図書館の中に置いてある機械でプリントし、製本して、売ってくれます。この場合、ディジタルコンテンツをキンドルで見ようが、ウェブで見ようが、冊子体に戻して見ようが一緒だよという例です。

先ほどディジタル化というのは、中身と入れ物との分離だと言いました。紙がいいと思う人は、そこで少し紙代というか印刷代を出して、それで物としての本を買えるんです。もちろん冊子体で買うほうが高いかもしれないけれども、そのほうがよければ、それに見合うコストをかければいいということになります。

そんなふうに考えていくと、例えば紙という物の場合、モノそのものにコストがかかるだけでなく、輸送や倉庫での保管など、物流にもコストがかかります。これに対して、ディジタルの場合、輸送コストはかかりません。オンデマンドで冊子体を作るコストが下がれば、必要に応じて、必要な場所で、利用者のニーズに合わせて利用のための形を決めることができます。何が言いたいかというと、ディジタル化が進んだことによって、我々のふだんの表現媒体というのはディジタルな環境の上に乗っかっていると思います。加えて、それを運ぶのもディジタルに運んでいます。読むときには、あるいは使うときには、いろいろなメディアを使います。それをうまく組み合わすことができれば、我々にとっては使いやすい環境にもできるんだろうと思います。

・ディジタルアーカイブ（Digital Archive）とは？
－ファジーな用語
・図書館、文書館、博物館や美術館のコミュニティで用いられている
・多種多様な情報資源、サービスがある
－ディジタル形式の情報資源を集め、蓄積・提供するサービスないし機能
・cf. Digital Libraries, Digital Curation
－一般にはそれなりに大規模な電子的情報資源のコレクションであり、主として歴史的、文化的コンテンツを扱っている
－長期にわたるサービスを前提とする
・なぜディジタルアーカイブを必要とするのか?
「文書や記録を将来に残すため」
・文書や記録を集め、保存し、提供する役割を持つ文書記録管理組織にとって、ディジタルアーカイブは重要な機能である
・私たちの情報環境、出版技術や出版スタイルの急速かつ根本的な変化によって重要さが増している
・ディジタル情報資源の長期保存は与えられた大きな課題

## ディジタルアーカイブ（Digital Archive）

ディジタルアーカイブという言葉についてです。アーカイブという言葉、公文書館ですとアーカイブなのかアーカイブズなのかによって、ざっと区別しないといけないのですが、ここではあまりきちんと区別せずにディジタルアーカイブと単数形で呼んでいます。ディジタルアーカイブと言う言葉はファジーな言葉で、結構実際に使うときというのはいろんな意味で使っています。基本的には、自分自身は、いろんなディジタルコンテンツを蓄積して長期にわたって提供していくサービスの意味で使っています。ディジタルライブラリというのもそうですし、あるいはディジタルキュレーション（Digital Curation）という言葉で呼んでいたりもします。とにかく長期にわたるサービスを前提とするというところが大事なところだと思います。

なぜディジタルアーカイブを必要とするのかについてです。文書や記録を将来に残すために、文書というものは昔からつくられてきました。ディジタルアーカイブでは、紙あるいは物としてつくられてきたものをディジタル化して、そして将来に残していくということもあれば、現在ディジタル形式でつくられるものをディジタルのまま将来に残していくということになると思います。そうした文書を集め、保存、提供するという役割を持つというのがディジタルアーカイブです。我々の情報環境が変わってきていますので、我々の情報環境の中ではごく基本的な機能であると思います。情報環境や出版のスタイルの変化の中で、重要さ増していると思います。その一方で、長期保存というのが非常にチャレンジングな課題です。

ディジタルアーカイブ－基本的視点
・メディアのタイプ
－パッケージ情報資源 vs. ネットワーク情報資源
－ファイルフォーマット

・作成プロセス
－物理的な資料からの電子化（Digitization, Turned Digital）
－もともとディジタル形式で作られたもの（Born Digital）

## ディジタルアーカイブの基本的視点

メディアのタイプは基本的な視点のひとつです。メディアのタイプとしては、パッケージ情報資源とか、あるいはネットワーク情報資源とかあります。ファイルフォーマットについても考えなければいけません。

ファイルフォーマットというのは、例えばこれはマイクロソフトの Word ですとか、あるいは Excel とか、あるいはそれ以外のいろんなもの、たとえば HTML のファイルとか、そうしたものですね。その作成プロセス、すなわち物理的な資料からの電子化、ディジタイゼーション（Digitization）ディジタリゼーション（Digitalization）、あるいはディジタイズド（Digitized）、あるいはディジタライズド（Digitalized）という言い方をします。最近聞いた言葉ですが、最初からディジタル形式で作られたもの

であることを意味する Born Digital（ボーンディジタル）に対して、ディジタル化して作ったものを意味するのにTurned Digital (ターンドディジタル）という言い方があります。ここではその言葉を利用しています。

ディジタル生まれという意味であるボーンディジタルの文書のほうは、ある意味で我々ふだんつくっている生のデータ、あるいは生の文書を意味します。保存という観点からすると、ボーンディジタルというのは、つくったときに使ったツールに依存しますので、そのツールそのものが本当に長く残っていくかどうかという不安があります。一方、ターンドディジタルのほうは、まあもともと長期にわたって保存するのであるという、その意図があれば、広く使われている標準を用いてディジタル化することによって、この問題を超えられるという強みがあります。

こうした点は、ごく一般的なことであると思いますが、ただ、自分自身は、余り違いはないかなと思っています。それは、ターンドディジタル資料の場合でも、例えば JPEG2000や TIFF 形式のように非常に広く使われているフォーマットで作って、それをそのままで持っていればいいんですが、でも実際にはその上にいろんな加工が施されていたりします。

加えて、JPEG2000にしても、あるいは TIFF にしろ、例えば本を１冊、あるいは何ページもあるものをディジタル化する場合があります。その場合、どのファイルが１ページ目、２ページ目、３ページ目ですといった付加的な情報を必ずつけないといけないです。ですから、ファイルのフォーマットだけが安定していますと言っていればよいかというと、やはり少し不足しています。ファイルに加えて、ファイルを使えるようにしている情報をきちんと保存していかないと、現実には使えなくなってしまいます。

そういう意味では、例えば PDF のファイルになっていると、何ページのものであろうが、それは１つの PDF ファイル、いわば PDF でつくられた本という単位で扱えますし、コンテンツも含めていろいろな情報も含めて生かすことができます。そうすると PDF、あるいはアーカイブ向けの PDF である PDF/A のように比較的安定している形式を選んでおけば、生のデータと自前でつくった、自分ところでしか通じないようなメタデータだけで持っているよりは安全であろうかと思います。

ボーンディジタルというのは、基本的にいろんな問題を含んでいます。でも、そうかといって、ボーンディジタルのままだと危ないから、いったんプリントアウトして、それをもう１回ディジタル化しようかといっても、なかなかそうは単純にはいかないと

思います。

・ネットワーク情報資源の収集
ー提供者とアーカイブ間の合意の上に行うもの
・組織内でのアーカイブ
ー収集ロボットを用いた自動収集によるもの
・Web アーカイブ

ネットワーク情報資源の収集の問題ですね。これについては、提供者とアーカイブ間の合意の上に行うものもあれば、あるいは収集ロボットを用いた自動収集によるものもあります。これは、下の場合ですと、必ずしもいわゆるウェブ上でのアーカイブだけとは限らないのですが、実際に文書を移管してもらうと、電子的なものを、例えばハードディスクに入れて、それを実際に物として持っていく場合もあります。それはそれなりに大変ですしコストもかかります。

では、ネットワークで送ればいいなとなりますが、ネットワーク上をどうやって安全に送っていくかが問題になります。基本的に文書を提供する側と、それからアーカイブする側との間での合意の上で仕事を進めなければなりません。その一方で、例えばウェブ上ではよく行われる、収集ロボットが勝手にウェブ上のコンテンツを収集してくるというものがあります。これは別にウェブの上だけに限らず、ネットワーク上でつながっている環境であれば同じことができます。

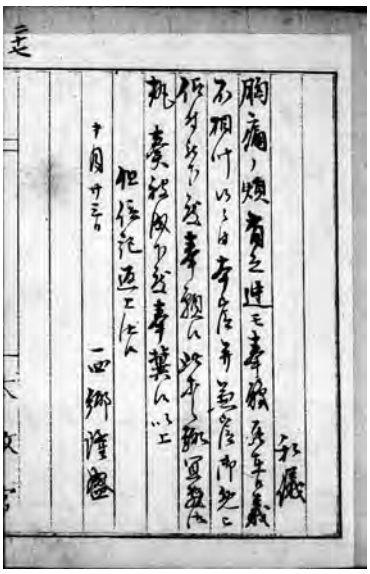
収集側と提供側の間で、何かの約束を決めておいて、この約束を守ってコンテンツを持っていってくださいねというようにすると、ある程度のコントロールは可能です。ただ、どういうものが持っていかれるかということを十分にコントロールしておかないと、余計なものまで持っていってしまわれるという危険は常にあります。ですから、収集というのも、提供側と収集側でうまく協働する、コラボレートすることを考えることが求められます。

・行政機関にかかわるディジタルアーカイブ
ーディジタル形式での記録の保存ーDigital Archive and Record Keeping
・電子的に作られ、利用された文書と記録を集め、保存する
ー記録の電子的提供
・インターネットを利用したリソースの提供-いつでも、どこからでも、誰にでも
ー歴史文書のミュージアムとしてのアーカイブ
・インターネットを利用して歴史文書を展示する
・学術および教育コミュニティに、高品質なリソースを提供する

次に、行政機関にかかわるディジタルアーカイブという話です。電子的に文書と記録を集め、保存することについてです。記録の電子的な提供、すなわち基本的には、インターネットを利用したリソースの提供です。「いつでも、どこでも、誰にでも」がディジタルアーカイブの特色です。こうした面を生かしていきたと思いますし、また、文化遺産としての歴史文書のアーカイブにもなっています。

例えば、アジア歴史資料センターのアーカイブがあります。明治維新から太平洋戦争終了の1945年までの間の外交文書、それから軍の文書と内閣の文書のとても大きなデータベースです。たしか現時点で1,900万イメージ超であったと思います。本当に世界最大級ですね。使われる機会も多く、我が国として本当に誇れるものだと思います。

ちなみに、私自身は歴史の研究者ではないので、こんなおもしろい内容があるよといって紹介する程度の非常に軽いユーザーで

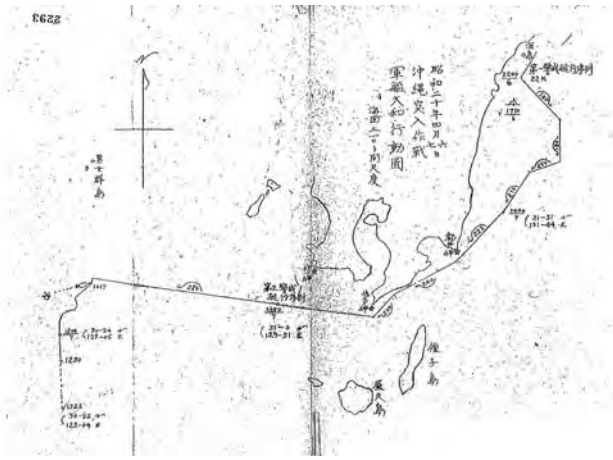


**西郷参議辞表**
**［レファレンスコード］A01100036300**

す。よく例として見せるのは、西郷隆盛の辞表です。明治維新の後に西南戦争があって、そのときに西郷隆盛が下野したという話は、多くの学生が知っています。でも、下野したことがこの辞表でわかるという文書の現物は見たことがないので、新鮮さを感じるようです。もちろんこれは写真ですが、現物の文書がそこにあるのであるということが良くわかります。内容が編集されて教科書に載っているというものではなく、自分で探し自分の目で見ることができるということは教材としてとても強力であると思います。

アジア歴史資料センターの場合、1,900万イメージの中でよいものを探していくというのは教師の役割であろうとは思います。その一方、教師がどこでも、いつでも使える素材を提供してくれるという点が、非常に強力なサービスである思います。

アジア歴史資料センターにいらっしゃった牟田さんの大きな努力で出来上がったものと思います。私自身、牟田さんからはいろいろ教えていただきました。生の歴史文書の持っている強力さというのは、ある種、ネットワーク経由で伝えることができるものであると思いますし、国として大事な歴史遺産として維持していかねばならないものであると思います。



**軍艦大和戦闘詳報**
**［レファレンスコード］C08030566400**

以前、牟田さんから教えてもらったことなのですが、有名な軍事史家から一般の人まで多くの人が知っている資料に、戦艦大和の最後の記録があります。撃沈されるまでの記録文書をディジタル化したものがあります。ディジタル化されたものではあっても、教科書に転記されたもののように２次的なデータとして編集されているものに比べて、本物の強みを持っていると思います。歴史を直接伝えるとでもいうのでしょうか。何かそういう意味での強力さというのは絶対ほかには負けないものがあるという感じがします。学術、教育コミュニティに高品質なリソースを提供しているというのは、本当にそのとおりであると感じています。

**一般化した視点**
**・文書のライフサイクルとアーカイブ**
**　ー作成、利用、保持、長期保存**

## アーカイブのモデル ― 一般化した視点

さて、一般化した視点として、基本的な点が幾つかあります。まず、文書のライフサイクルとアーカイブ、作成、利用、保持、長期保存、こうしたことを考えます。これは例えば文書のアーカイブを考える場合、保存のところだけを考えているのではなく、

生まれたところから最後の保存のところまで全部通して見ていかないといけないと思います。

・アーカイブの機能
　－収集、保存、提供
　－メタデータ：記述、管理、技術的

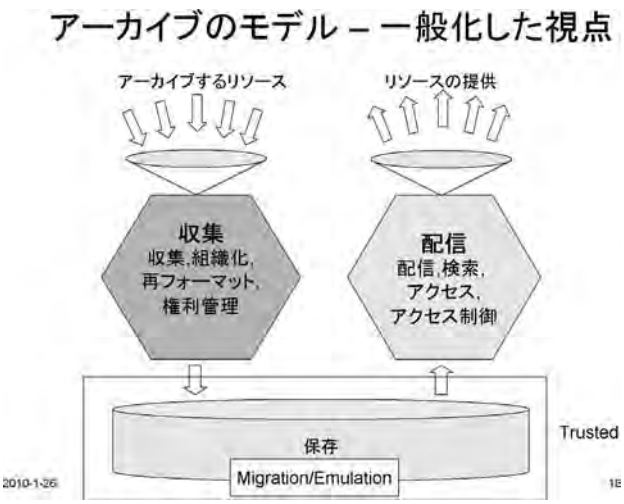
それからアーカイブ機能の部分だけ取り出してみると、収集、保存、提供ということになると思います。その収集、保存、提供の上で必要なメタデータとしては、文書の内容の記述、すなわちどういうコンテンツであるということの記述をする。それから、どのようにして管理していくかについての記述。たとえば、だれが見てもいいのか、利用者が見る前には必ずチェックしなければいけないといったこと、どういうファイルフォーマットなのか、いつつくられたのか、いつそのファイルの形式変換されたのかといった技術的な内容です。そうした記述が必要です。

・アーカイブ方針と戦略
　－保存のためのリソース選択と組織化
　　・どんな実体を保存すべきか？
　　・どのような機能を保存すべきか？
　－保存のための効率的なリソースの加工
　－保存のための効率的なメタデータ作成

それから、アーカイブの方針と戦略、保存のためのリソースを選択して組織化するということの記述。アーカイブの方針を立てないといけないですし、その方針に基づいて戦略を練っていかなければいけません。それから、保存のための効率的なリソースの加工も必要ですし、それから保存のための、これが結構面倒くさいのですが、効率的なメタデータの作成というのもあります。

ディジタルアーカイブを前提に、実際に保存しアーカイブするということを考えると、紙のものであればそれをディジタル化するときには、それ以前に、何を保存するのかという選択があるでしょう。さらに、資料が持つどのような機能を保存すべきかも考えなければなりません。紙の資料を電子化すると、ぱらぱらめくることができるといった使い勝手の良さが失われることになります。紙の資料が持つ機能を失っても、電子化することが良いのか、電子化する際にテキスト検索のような新しい機能を付加するべきなのか、といったことを考えなければなりません。

次に、電子的につくられたものをアーカイブする場合、例えば動画が入っていたけれども、この動画というのはどうも残すのは難しそうだ、では動画はやめてしまおうとか、あるいは静止画、スナップショットでとっておけばよいといった判断を必要とします。同じように、ハイパーリンクの保存は難しいのでリンクの保存は考えないとか、あるいはそのハイパーリンクに関する情報だけは残しておくけれども、リンクそのものを保存しないといったことを決め、その決定に従って加工していかないといけません。




これは、一般化した視点を図であらわしただけです。左側に入ってきて、最初は収集です。

ボーンディジタルなものをそのままの形で残しておくこともあれば、保存に都合のいい形、例えばワードの文書ですね、PDF/Aの形に変換するといった場合もあります。また、この時点で権利管理のためにいろいろな情報を取り出しておかないといけない場合もあると考えています。

そして保存、それから配信ですね。配信の際には検索、あるいはアクセス制御等も行いますが、まず保存のところですね。技術はどんどん変わっていきますので、新しい技術環境の中に移植していくというか、移住させていくという方法があります。これをマイグレーションと言います。それから、新しい環境の中で古い環境と同じ環境をつくり上げる方法があります。これをエミュレーションと言います。こうしたものが代表的な保存方法として考えられてきています。

いずれにせよ、ここのところで大事なことというのは、いろいろなフォーマットでつくられているリソースが作られているということと、それをとにかく長生きさせていかないといけないということです。現在、既にビットデータとして、あるいは非常に限られた種類のファイルフォーマットのファイルであれば、長期の保存を保証しますというアーカイブサービス、あるいは保存サービスをしているところがあります。ただ、基本的にビットデータの保存であれば、それ自身はそんなに難しくないと思います。

だけど、現実にはいろんな種類のソフトやツールを使っていますので、そのソフトやツールを使ってつくられたファイルをもとのまま、例えば30年後まで残しておくということ、それ自身はやはり難しいです。ですから、どんどん新しい環境に引っ越していかないといけない。あるいは、古い環境をまた新しい環境の上に実現していくということをしていかないといけないと思います。

自分自身は、基本的には100%そのままの形ですべての電子的な文書が持っている機能を残していくというのは、やはり難しいだろうなと思います。ですから、そういう意味では残す部分、すなわち「これだけの機能が残ればいいよ」ということを決めなければならないと思います。例えばテキストというのは、普通のWordの文書の中身はテキスト検索できますね。これは、ある種、電子文書として大事な機能です。ところが、例えば「人間が読めればいいよ」ということが保存条件であれば、Wordの文書そのままで保存することをあきらめて、ページイメージデータにしてしまうという方法も可能性としてはあります。

ただ、そうしてしまうと検索性が全くなくなるので、現実問題として使い物にならないと考えるのであれば、例えばＰＤＦのアー

カイブフォーマット、PDF/A を選べばよいわけです。これは国際標準規格でもあり、長きに渡って安定するでしょうから、保存の目的に合わせて適切な形式を選ぶほうがいい。ですから、保存のための完璧な方法を作っていくというよりは、保存のためのガイドラインを作っていくことのほうが大事だろうなと思っています。

図の Trusted と書いた部分というのは、信頼できるものでないといけないという意味です。さっき言った保存のサービスをしているというところは、この部分のサービスをします。アーカイブを全体で考えると、収集から保存、配信まで全部なのですが、データだけを預かって保存しますという巨大な安定したデータセンター的サービスもあります。高い信頼性を保ってディジタルデータを保存することに非常に大きなコストがかかるということであれば、いろいろな組織が保存機能を共有して持つことは当然のことであると思います。当然共有するほうが、コストパフォーマンスはよくなると思います。

**課題**
- **収集と保持の基準**
- **リソース組織化の方針**
- **保存のためのリソースの再フォーマット－安定したフォーマット、文書と記録の一貫性管理**
- **権利管理－著作権、個人情報等**
- **リソースの内容記述と保存のための記述にかかわるメタデータ**

## アーカイブのモデル－収集フェーズでの課題

課題として、収集と保存の基準をつくること、あるいはリソースの組織化の方針ですとか、あるいは保存のためのフォーマッティング、あるいは再フォーマッティングですね、そうしたものを考えなければいけないとかといったことが課題になります。スライドの一番下に出てくるメタデータというのは大事な要素です。ただ、人手でメタデータを書くと非常にコストがかかるので、できるだけ機械的につくることですとか、もともとつくられていたメタデータを取り込むというようなことをしないといけないです。

アーカイブというのは文書のライフサイクル全体で考えないといけないです。文書は、最初つくられたところできちんとその文書に関する情報をつくっておいてもらえば、後でその情報をコピーして使えるケースは多いはずなんです。ただ、もとのところでちゃんとつくってくれていないと、保存をする人がもう１回つくり直さないといけないことになりますから、それはやはり無駄ということになります。

**保存フェーズでの課題**
- **信頼できるアーカイブの維持コスト。ことに小さい組織にとって－組織の改変に対処しなければならない**

## アーカイブのモデル－保存フェーズでの課題

保存についての課題ですね。とにかく信頼できるアーカイブの維持コストが課題かなと思います。特に小さい組織にとっては、そのコストに耐え得るかどうか。

提供フェーズでの課題
・検索とアクセスの機能
－メタデータによる検索、テキスト検索、イメージ検索など
－障害を持つ利用者のための Accessibility に関する課題

## アーカイブのモデル－提供フェーズでの課題

次にリソースの提供。リソースの検索とアクセスを提供します。このときに、一般的なことですけれども、メタデータによる検索、テキストの検索、イメージ検索、いろんな検索機能をつくっていかないといけないでしょうし、加えて、障害を持つ利用者のためのアクセシビリティーに関する課題を考えておかないといけないです。例えば、視覚に障害を持つ利用者、あるいは手が動かせない利用者、そこにはいろんな利用者があります。これも情報システムというか、情報を提供するシステムでごく一般的な形です。

・アクセス権限の管理
－利用者による管理、利用場所による管理など
・原本性の保証
・複製の製作

それから、プライバシーの問題などでアクセス権限の管理をすることも当然必要です。それから、また原本性の保証といった課題もあります。原本性の保障のために認証システムを利用するという話題が良く出ます。ただ、私自身にはちょっとよくわからん部分もあるんです。それは、常にこういう配信サービスが24時間365日提供されていて、加えて、保存サービスが本当に信じられるものであるならば、いつでも利用者は適切なコピーがもらえることになります。すると、常にダウンロードしてきたコピーは信じることのできるオリジナルのコピーですね。それと自分が持っているもの間の比較というのはいつでもできることになります。そうすると、難しい認証システムが本当に要るのかなと疑問に思うんです。

余り難しく考えなくても、信頼できる保存システムがあれば原本性というのは常に保証されているのではないかなと思ったりするんですが、ここのところはよくわかりません。でも、いずれにせよ公文書を提供するわけですので、オリジナルの公文書ですよということを何らかの形で保証しないといけないと思います。それは、研究されるべきテーマです。ただ、その時点で生きている公文書とは呼べないですので、例えばいわゆる文書の認証システムのようなものを使わないといけないかどうかについてはよくわかりません。

収集フェーズでの課題
・安定したフォーマットでの保存
－安定性と引き換えに文書が持つ何らかの機能を失うことになる。
－文書の一部の機能を犠牲にすることに関する了解と合意の必要性

安定したフォーマットでの保存について、先ほど、例えばWord ファイルをイメージデータに変えても良いのではといったことを言いました。電子文書に限らず、文書は何らかの機能を持っていますが、その機能を落としながらでも安定したフォーマットに変換して保存することを考える必要があります。

この文書機能を落とすことについて、電子文書にある意味で話を特化していますが、随分前に、ミシガン大学のマーガレット・ヘッドストロムさんから聞いた話でもあります。彼女は90年代、最初にディジタルアーカイブが出てきたころから保存のことをちゃ

んと考えなければならないと主張してきた方です。もともとアーキビストですけれども、彼女から、電子保存するときには、ある程度その機能を限定するようなことを考えないといけないけれども、それは電子保存に限った問題ではなくて、もともと紙のものをマイクロフィルムにしていたときにだって、冊子体の使いやすさという機能を失う代わりに、コンパクトで安定した保存を行うことを選んできたのであるということ聞いたことがあります。

冊子体を置いておくことにもいろいろな問題があります。たとえば、場所の問題、劣化の問題などです。本当は紙で残したいのだけれど、中身をみるのにはマイクロフィルムで代替できるからマイクロフィルム化して残している、といったところであると思います。これまでもいろいろな制約条件の下で現物を廃棄して中身を残すということをしてきたわけです。ですので、ディジタルになっても同じことなんだということをヘッドストロム先生から聞き、当たり前ではあるのですが、なるほどなぁと感じたことをよく覚えています。そうすると、残すべき中身とは何なのかということがきちっと決まっていればいいし、それを決めることというのが、その保存の中心課題になるということになるのかなと考えています。

**・「そのまま」保存**
**ーオリジナル版をそのまま残す**
**ー再現性を失う危険性**

場合によると、やはりオリジナルをそのまま残さないといけないものも当然あると思います。ですから、それはケース・バイ・ケースであって、それぞれのケースについて判断しやすくするためのガイドラインを作っていかねばならないのだろうと思います。

**何を保存するのか？**
**・一般に、たとえ「望ましい」とは言っても、原フォーマットのままディジタルリソースを保存するにはコストが高くつく**
**・これに対するひとつの解決策は、原フォーマットに比べて安定していて保存しやすいフォーマットに変換して保存することである。たとえば、ページイメージや PDF のような印刷用フォーマットのファイル。**
**・変換によって何らかの機能を失うことになる。たとえば、ハイパーリンクやテキスト検索の機能、そして look-and-feel は失いがちである。**

次に、何を保存するのかの話題です。これは、今言ったようなことですね。一番下にこういうふうに書いていますが、これはボーンディジタルなものは特にあてはまるのですけれども、保存のための変換によって何らかの機能を失うことになります。だから、どこまで何を残さなければならないかということです。

ルック・アンド・フィールというのは、要は使い勝手ですね、見ばえとか手ざわり、そうしたものまで残さないといけないとなると、それはもうマシンそのものから残していかないといけないです。例えば、今使っている Wii のソフトの保存を考えてみましょう。30年後に Wii がそのまま残っているかは不明ですが、その一方、Wii リモコンによる使い勝手を含めて残さないといけないということになると、そのマシンそのものを残さないといけないことになります。でも、マシンそのものが残せるかどうかわからないですね。それから、私の子供から聞いたのですが、前に持っていたゲームキューブのソフトが Wii にも使えるようになっています。使い勝手は異なりますが、息子は何も不満に思ってい

ないようです。このように　使い勝手が変わっても、要は満足できるのであれば、それはそれで保存できていることになっています。

**メタデータ**
**・データに関する（構造化された）データ**
**（Structured）Data about Data**
**・記述対象に関する「何か」を書いたもの**

## 2.　メタデータ

ではメタデータです。メタデータとは何か。多分、メタデータという言葉は今までどこかで聞かれたことはあると思います。私自身、メタデータという言葉については、95、6 年頃に聞いたときは、あれっという感じがしたんです。それは、それ以前から知っていたコンピューター関連の知識としてのメタデータとは少し違ったからでした。

**メタデータの課題**
**・メタデータはディジタルリソースのアーカイブにとっての重要な要素**
**・メタデータにかかわるコストの削減**
**－人手による記述を前提とする内容記述のメタデータは、文書のライフサイクルの前の段階で書いてほしい**
**－技術的、構造的メタデータの自動抽出**
**－メタデータ作成と管理のためのソフトウェアツールの必要性**

それではじめはよくわからなかったんです。メタデータというのは、データに関する、何らかの情報資源に関するデータ、あるいは書いたものです。私は、メタデータの説明のときに、いつもこういうペットボトルを持ってきます。それで何をするかというと、ボトルについているラベルをはがします。さっきまではお茶のボトルでしたよね。こうなったときにこのボトルは何でしょうか。さっきの状況を見ずに、これをどうぞといってわたされても、中身が何であるかわからなければ気持ちが悪いですね。まず飲めないです。相手が信じられる相手だったらば飲んでもいいかなと思うかもしれないですけれども。では、ペットボトルが自動販売機から出てきてなぜ飲めるのか。ラベルにいろいろな情報が書いてあるからです。この情報を頼りにしています。では、このペットボトルをウェブページだとか、あるいはウェブ上に置いてある文書ファイルと思ってください。これをどんと渡されて、これに関する目録をつくってくださいと言われたとします。すると、中身がわからないなりに書くしかありません。言ってみれば、中身を飲んでみるしかないわけです。

もちろんこれをつくっている人とか、つくっていたところを知っているような人、あるいはどうやって入手してきたかということを知っている人であれば、これに関する情報を書けます。しかし、そうでなければ、すべて自分で中身を確かめて書かなければいけないことになります。先ほど文書の保存ということを話した際に、アーカイブでは、文書のライフサイクル全体で考えないといけないですよと言いました。保存をする場に文書を持ってくるのに、裸のペットボトルのような形で持ってくることはよくありません。

メタデータを考えるときに、ペットボトルの例をよく使います。実際にこの場合、中身はこのボトルの中に入っている液体です。液体ですから、いかようにも形は変わります。そういう意味では、

ディジタルコンテンツと本当に似たようなものなんです。蛇口から液体が出てくるのと同じように、ネット経由でコンテンツがやってくると考えれば良いのです。すると、それに対して、どういう内容のものであるか、どういうつくり方をされているのか、どういう性質なのかということもやはりきちっと書いておかないといけないし、それを保存する側、あるいは受け取る側の立場からすると、つくったところで書いておいてもらえると本当に助かるということが理解できます。

それで、メタデータに関してなんですけれども、メタデータは「データに関するデータ」と言うだけで終わりになってしまいます。けれども、ネット上で、我々が何かを探して、そこにアクセスして、サービス受けながら評価します。例えば、自分でどこかに旅行に行くときのことを考えてみるとします。以前だと、とにかく旅行屋さんの案内を見て、いろんな情報を得ましたし、それから店頭でいろんなことを教えてもらったと思います。だけど、今結構ネットで旅行の情報を仕入れて予約もしてしまいますね。どの宿がいいかを調べるのにレビューを見たりとか、それからどこが安いとか、例えば価格を比較したりとか、いろんなことを自分でやっています。予約をする、あるいは航空券を買うというふうにいろいろなことをします。クレジットカード情報も渡してしまうわけですね。ですから、相手がどういうものであるかということをきちんと信頼できないと、そこまではできないはずです。

では、そういうとにかくいろいろなことを、ネット上でしているんです。けれども、決して今目の前にいる人に頼むのではなくて、ネット上のどこかにつながっているサービスに対して頼んでいることになります。アクセスしているサイトが東京にあるのか、あるいは香港にあるのかわからないです。そうした環境で、我々は何の気なしにサービスを利用しているんです。

では、そこで何を信じてサービスを使っているかというと、そこで提供される情報が要は信じ得ると考えているわけです。ですから、そういう意味でも、いろんな作業をする際に、そのサイトであったり、あるいはそこでいうサービスであったり、それに関するいろんな情報を使って、すなわちメタデータですね、それを使って我々はネットを使っていると思います。

- **目録や索引なしに図書館や公文書館のサービスは考えられない。目録や索引は典型的なメタデータ**
- **ネットワーク時代におけるメタデータの役割**
  - **ー探す、選ぶ、アクセスする、利用する、管理する、保存する、そのほか**
  - **ーこれをネットワーク越しに行うとすれば、メタデータの重要性が直感的に理解できる**

## メタデーターどのようなものがあるか

ネットワークの世界でメタデータの話というのは、探すだけだったらグーグルでもいいかもしれないのですが、そこから先のことになると、いわゆるちゃんとしたメタデータがないとやっていけ

**どのようなものがあるか**

- **目録、索引、抄録**
  - **ー資源を探すことを目的に、資源の内容を抽出して記述する**
  - **ーどのようなものがあるか、どのように利用するか？**
    - **OPAC、テレビ番組、オンラインショッピング**
- **辞書、事典、シソーラス**
  - **ーことば・概念について書いたもの**
- **識別子**
  - **ー資源につけられた「名前」**
  - **ー「何か」を識別し、指示するもの。**
- **権利（権利管理）**
  - **ー資源に関連するいろいろな権利・権限**
  - **ー使う権利、見る権利、改変する権利、etc.**

・利用者（利用者対象）
　－大人、子供、何らかの障害を持つ利用者
・視覚、言語（ことばのわかりやすさも）
　－Rating
　　・18歳以上、15歳以下は親（大人）と一緒に見ること、etc.
　　・性的、暴力的内容
・利用方法
　－資源の使い方
　　・使い方、適用年齢、関連資料
　　・資料の見方（読み方、利用方法）の説明
・利用環境
　－資源を利用するために必要な環境
　－資源
・イベント
　－何らかのできごとに関する記述
　　・出版したこと、翻訳したこと、etc.
　－「もの」というよりは、「こと」に関する記述
・その他

ない。では、メタデータにはどんなものがあるかというと、目録とか索引、辞書とか事典とか識別子、こういうのはごく一般的なもので、権利管理、あるいは利用者対象について書いたものがあります。

あるリソースを利用者に提供するときに、例えば視覚に障害がある利用者もいますので、提供の仕方を変えないといけないです。そのため、利用者の特性に関する情報と、その提供する内容に関する情報、あるいは提供に使う道具に関する情報、それをマッチングしたりしないといけないです。だから、いろんな情報を我々は使わないといけないです。それで、利用方法、利用環境、イベント、その他、とにかくいろんなものがあります。

いろんなものがありますので、そのメタデータの国際標準、標準規格というのは、これまたいっぱいあります。ダブリンコア。MARC は図書館資料対象ですし、ダブリンコアはインターネット全般です。それからディジタルアーカイブのコンテンツ関連では、PREMIS、EAD、METS、公文書館の関係ですと ISAD（G）といったものがあります。それから保存ですと OAIS の参照モデルです。それから教育・学習資料に使う LOM ですとか、政府情報ですと GILS、ネットワーク指向になっているオーストラリアの AGLS、こうしたものがあります。ビデオですと MPEG 7というものがあります。このように、とにかくいっぱいあります。それで幾つかとにかく名前だけでも出してみました。

・ディジタルリソースのアーカイブで用いられるメタデータの国際標準
　－ISAD(G)、EAD、METS、OAIS、PREMIS、…

## MODS

MODS：Metadata Object Description Schema
・MARC21から選び出したエレメントによるメタデータの記述
　－簡略化
　－MARCXML：MARC 21の XML 記述
・XML による記述

MODS、これはメタデータ・オブジェクト・ディスクリプション・スキーマです。広く利用されている MARC というのは非常にたくさんの記述項目を持っている標準規格です。

巨大な規格というのは、限定されたコミュニティの中で限定された資料を扱う上では都合がいいですけれども、ネット上のようにいろんな種類のものを扱わないといけないときに、余り厳密にがちがちに決めても、なかなか例外的なもの、多種多様なものは扱いにくくなってしまう。そうすると、かえってシンプルにするほうがいいよということになってきます。MODS はこの MARC をベースにしてかなりシンプルにしています。あとは、ネット上での流通を考えますので、XML を必ず考えます。

## AGLS

AGLS
・オーストラリアの政府行政情報のためのメタデータ
・GILS：Government/Global Information Locator Service
　－行政情報のためのメタデータ
・AGLS は Dublin Core を基礎
　－拡張
・オーストラリア NAA：National Archives of Australia
・USA：NARA National Archives and Records Administration

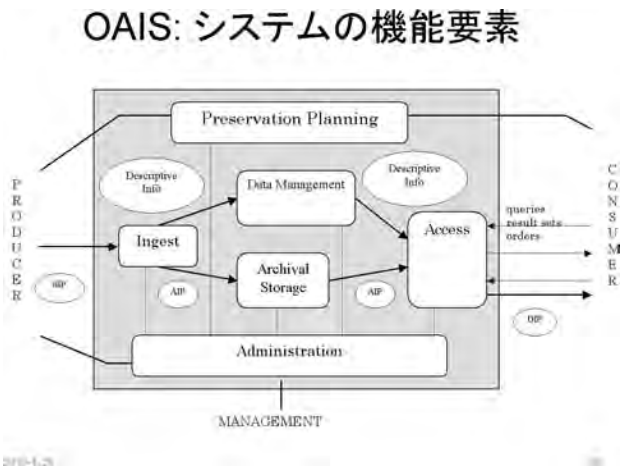
これも90年代からやっているものです。オーストラリア政府がネット上での情報アクセスのためにつくっているもので、文書と

かサービスを提供するときに、そうしたリソースに関する内容を記述するためのものです。そして、その内容の記述を使ってウェブ上でリソースを探すためにつくられているものです。

**OAIS**
- **OAIS：Open Archival Information System**
- **OAIS の参照モデル**
  - **ーアーカイブシステムの要素**
  - **ー情報オブジェクトの構造**
    - **ビットデータ＋再生のための情報**
  - **ー情報パッケージ**
    - **情報オブジェクトを入れるいれもの**
    - **情報オブジェクトに関する保存情報**

## メタデータの基本モデル例－OAIS

それから OAIS、オープン・アーカイバル・インフォメーション・システムです。これ自体は、保存システムの参照モデルです。ここでは、余りその中身を詳しくは説明していなくて、ごく簡単に書いてあります。プリザベーションプランニングだとかアドミニステーションというのは、方針決めとか全体の管理に関する、どちらかというと人的な要素です。要するに、このモデルは人と機械を含む全体のシステムになっていますので。一方、真ん中にある要素は機械的な部分です。アーカイバルストレージというのは、いわゆるデータベース本体ですね。それからデータマネジメントというのは、その内容を管理するものになります。それの左右にインジェスト、いわゆるリソースの取り込みと、それから利用者（コンシューマ）側からのアクセスを受け付ける要素があります。




このモデル自体は、ただそれだけの話です。ここに小さく DIP とか、真ん中に AIP、あるいは左側に SIP というのがあります（これはインフォメーションパッケージで IP と呼びます)。サブミッション・インフォメーションパッケージ（SIP)、それからアーカイバル・インフォメーションパッケージ（AIP)、それからディセミネーション・インフォメーションパッケージ（DIP）という名前です。

インフォメーションパッケージというのは何かというと、パッケージですので、簡単には段ボール箱だと思っていただければいいんです。それで、その段ボール箱を要はどこかに、倉庫に入れておくわけです。、箱に物を詰めて入れておくことになります。その箱の中に入っているものはというと、左側がいわゆるコンテンツなんですね。それで右側は、そのコンテンツについて、今まで保存してきたいろんな来歴に関する情報ですとか、あるいは途中で行った加工に関する情報ですとか、そのコンテンツを保存していく上で必要な情報を、このパッケージの中に、すなわち段ボール箱の中に一緒に入れておくんです。

それで、ふたをしてしまうわけですけれども、ふたをして外側

に何も書いておかないと、その箱そのものをあけないと中身がわかりません。ですから、それでは困るというので、そのパッケージに関する記述、中にこんなものが入っているというのをぺたっと張っておくわけです。これがこのインフォメーションパッケージのイメージです。

インフォメーションパッケージの中身ですけれども、コンテントインフォメーションというのが、実はもともとビットデータです。データオブジェクトというのはビットデータ、ビットの並びですけれども、それだけだとやはりコンピューター上で利用できませんので、ある種コンテンツを使うための情報を足しておかないといけないです。例えば Windows なんかでも、ファイルの拡張子を持っていて、それを使ってファイルの種類を解釈し、適切なソフトウェア起動し、中身を再生します。

このようにコンテンツを使うための情報を足しておかないといけないです。例えば、これは Word の文書だよとか、使うのにはどういうツールが必要だよとかいう情報を含めて、それで一つの使い得る情報オブジェというふうになります。ですから、ビットデータだけが残されていたのでは、これはもうどうしようもありませんので、それを使うため、再生をするため、再現をするため、あるいはそのプレゼンテーション、すなわち表現をするための道具に関する情報をくっつけて、それを箱の中に入れて、そしてさっきの保存システムに格納する、あるいはそこから取り出すことになります。こうしたことを国際標準として決めています。

**METS：Metadata Encoding & Transmission Standard**
- **Digital Library の中の実体に関する、記述的（descriptive）、管理的(administrative)、構造（structural)的なメタデータ**
- **OAIS（Open Archival Information System）に対応**

## METS

国際標準としての OAIS モデルをベースにして決められているものの一つが METS と呼ばれるものです。METS は７つセクションからなります。結構複雑ですね。実際には機械的につくり出す部分が多いです。国立公文書館で行われていますディジタル文書の保存のシステムでも METS をベースにしたシステムを今検討されていると思います。

**PREMIS のモデル**
- **PREMIS：ディジタルリソースの保存のためのメタデータ**
  **ー http://www.loc.gov/standards/premis/**
- **知的実体とディジタルオブジェクトを分ける**

## PREMIS

PREMIS というのがあります。これもメタデータの標準です。PREMIS は、アメリカの議会図書館ですとか、OCLC、そうしたところが中心になってディジタルリソースの保存ということのためにつくったモデルです。記述項目を決めるデータディクショナリーがつくられています。

このモデルで、おもしろいのは、図の左側を見ていただくと、

ディジタルオブジェクトと、それからインテレクチュアルエンティティーに分けています。これは、「同じ文書なんだけども、ワードでつくられているものと PDF になっているものとありますよ。そうすると、中身は一緒だけども実体としては複数あるんだよ」というケースを表現するためのものです。それをこのようにで分けているんです。

ディジタルオブジェクトとして、例えばこのパワーポイントのファイルですけれども、皆さんにお配りするのに PDF に変換しています。ちょっと形式は異なりますが、中身は一緒ですよね。文書としての機能は異なりますが、PowerPoint 文書と PDF 文書が持つ知的内容（インテレクチュアルエンティティー）は、ここで講演のタイトルを表すようなものですし、ディジタルオブジェクトでは２種類になる。さらに、それに関連する権利ですとか、あるいはこういう events というのは、何から何がつくられたといった、文書に関連する何らかの事象を表します。このように保存のために、ここに５つの要素を決めています。

## ダブリンコア

ダブリンコア
- Dublin Core はインターネット上でもっともよく知られている（使われている）メタデータ
- インターネット上の多種多様な情報資源の発見のためのメタデータ
  - ー Dublin Core Metadata Element Set（DCMES）
  - → 多様な資源に共通な記述要素
  - → 1995年ごろから
- Descriptive Metadata
  - ー情報資源の発見のために有用な属性の記述
- 草の根的参加者による開発
  - ーメーリングリストでの議論とワークショップでの合意形成
  - ー Dublin Core Metadata Initiative（DCMI）
    開発と維持管理のための組織
- Simple Dublin Core（15エレメント）
  - ーISO 規格：ISO 15836
  - ー日本の規格：JIS X 0836 ダブリンコアメタデータ基本記述要素集合

さて、最後、ダブリンコアです。ダブリンコアという名前、結構聞かれると思います。メタデータという話、特にインターネット上でのメタデータという場合にダブリンコアは結構出てきます。全く聞かれなくても、別に何の不思議もないんですが、ただ、よく使われます。

よく使われている一つの理由についてです。もう15年ほど前からつくり出してきたものなんですけれども、シンプルダブリンコアという名前で、15のエレメントでメタデータを書こうよというものが、最初有名になりました。それで、この後のスライド、ここに15のエレメントが書いてあります。ダブリンコアがよく使われている理由というのは、こういう基本的なエレメントを決めているところにあります。

普通、目録のためのデータベースをつくると考えると、どういう記述項目を用意するかということの記述項目ごとに、この項目は必須ですとか、この項目は省略しても構いませんと決めますよね。さらに、その記述項目ごとに、さらにその３項目みたいなことを決めていくことが多いですよね。それから、その実際の項目の値として書くならこういう形式で書くということを決めます。

普通はそうした規則をきちんと決めます。そうしてきちんと決めることは、その目的、ある一つの応用の中では当然必要なことです。一方、別の組織との間でデータをやりとりしよう、いわゆ

**ダブリンコアの基本15エレメント**

| | | |
|---|---|---|
| タイトル | Title | 情報資源に与えられた名前 |
| 作成者 | Creator | 情報資源の内容の作成に主たる責任を持つ実体 |
| キーワード | Subject | 情報資源の内容のトピック |
| 内容記述 | Description | 情報資源の内容の記述 |
| 公開者 | Publisher | 情報資源を利用可能にすることに対して責任を持つ実体 |
| 寄与者 | Contributor | 情報資源の内容への寄与に対して責任を持つ実体 |
| 日付 | Date | 情報資源のライフサイクルにおける何らかの事象に対して関連付けられた日付 |
| 資源タイプ | Type | 情報資源の内容の性質もしくはジャンル |
| 記録形式 | Format | 物理的表現形式ないしディジタル形式での表現形式 |
| 資源識別子 | Identifier | 与えられた環境において一意に定まる情報資源に対する参照 |
| 出処 | Source | 現在の情報資源が作り出される源になった情報資源への参照 |
| 言語 | Language | 当該情報資源の内容の言語 |
| 関係 | Relation | 関連情報資源への参照 |
| 時空間範囲 | Coverage | 情報資源の内容が表す範囲あるいは領域 |
| 権利管理 | Rights | 情報資源に含まれる、ないしは関わる権利に関する情報 |

る相互運用をしようとしたときに、同じ規則でもって同じ形式のデータを扱っている組織の間であればきちんと決まっている規則がありがたいです。これは、お互いに同じデータを使っているので。ところが、ちょっとでも違ったデータをつくろうとすると、そこで突然相互にデータの交換ができないという問題が生じてきます。

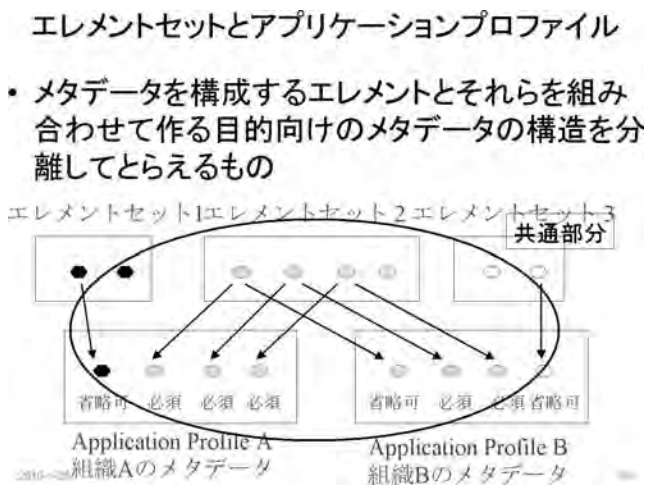
では、インターネットの世界ってどんなものかというと、あっちこっち違うものがある世界です。違うものがあって、その違うものの間で、あるいは違う組織の間でデータを交換したいよねというところです。例えば、資料館と図書館の間だと、例えば同じ県にあっても多分データのつくりが随分違うと思います。では、その間でデータをやりとりしようよと考えたときに、そうすると、合わせられるところだけ合わせようよというふうにおのずとなっていくと思います。そうすると、合わせられるところだけ合わせようよといっても、そこを表現するための記述項目の名前を合わせるところで結構問題になると思います。

では、そのときに、もしインターネット上で標準的に使われている記述項目というのがあれば、その記述項目を使いましょうねとすればどうでしょうか。こうすれば自分たちのものを押しつけるということにはなりません。加えて、ある県の資料館と図書館の間だけでやりとりをするという話がもし広がったとすると、国内全体、あるいは世界でつながっているということになるとすると、どこに出ても使える記述項目を使っていると相互運用が進めやすいですね。

実際にダブリンコアというのが目指したことというのは、共通に使える記述項目だけを決めることなんです。加えて、そういう共通の記述項目をベースにして、自分たちの目的であるメタデータの記述規則を決めるための考え方、それを提案してきたんです。

その考え方というのは、実はスライドの後ろのほうにあるんですけれども。例えばこういう左側にある一つの目録とかメタデータの規則、それで右側にある別のメタデータの規則があります。ここで記述項目の定義をそれぞれの中で閉じてしまうと隣同士でうまくやりとりできないんです。けれども、記述項目の定義はどこか別のところで、世界のどこかで定義してくれたものを借りてきて、それの意味を変えずに使うとすれば、すなわち、そうした記述項目を、そういう名前で、かつその意味でもって使いますよということにしておけば、例えばこのＡとＢのところでは、真ん中の⬭の部分は記述項目をお互いに共有できることになります。

こんな感じで、項目立てをしていって、規則をつくっていけば




いいよという、そういうことを提案してきました。これをダブリンコアのアプリケーションプロファイルと呼んでいます。

ここに15の項目があります。この15の項目、15という数は別にマジックナンバーでも何でもありません。たまたま15になったという、それだけなんです。実際に、この15というのは、減らそうという議論がありました。そういうものだと理解してもらえば良いと思います。ここに出ているものを見ていただきますと、タイトルとか、作成者とか、キーワードとか、内容記述とか、大体どこででも使えそうなものです。ある種、ダブリンコアにとっての一つの「肝」の部分という意味では、例えばここに作成者という言葉を使っています。決して著者とか作曲者とか、そういう名前は使っていないです。できるだけ一般化できる名前を使っています。

実際にはこういう定義に基づいて作られたメタデータをやり取りすることになります。データ交換のためには記述項目の意味やメタデータの表現形式は統一されていなければなりません。その一方、ネット上で実際にデータを交換するためのときの形式と、利用者が直接目にするときの形式というのは一緒である必要はありません。その一方、利用者にとってはわかりやすい名前を使う必要があります。例えば、図書館でもってダブリンコアを使うというときに、例えば作成者を著者にしてもそれは別に構わないです。データベースをつくるときだけ、どこかで著者と言い、あるいはどこかで漫画家と言い、一方でこれらは同じだよということがあってもかまいません。データベースの上で矛盾がしなければいいわけです。そういうふうにして、共通に使えるものというのを、最終的に15項目提案したものが、ダブリンコアの最初の要素集合です。

でも実際、この15を見てみるとアバウトなところに気がつくと思います。例えば Date エレメントは日付という意味だけです。多分このままでは使いにくいですね。例えば文書をつくった日付とか、文書が有効でなくなる日付だとか、そうしたいですね。でも、最初の15のエレメントではここまでしか決めなかったんです。その後、より詳しい意味を持つエレメントを決めています。

ここに Creator（作成者）とか、Contributor（寄与者）とか Publisher（出版者）とかあります。これらの厳密な区別というのは実は難しいのです。

- **DCMI が認定しているエレメント（記述項目）は71ある**
- **Simple DC 自体は国際標準として生きている**

実際に今までにこれらを１個にまとめてしまおうかという議論がありました。これらをまとめてしまっているようなエレメントとしてはつくっていないんですが、そういう概念というのも含まれています。それで、現在は70ぐらいになっています。

**・Dublin Core 自体の意義は、 Web 上でのメタデータの相互運用性に関する概念をはっきりさせたこと**
**ー記述項目と構造定義の分離**

ダブリンコアというのは、ネット上でいろんな違う組織でつくられたデータをお互いに流通をするための基本的な考え方を提案してきたものです。このエレメントというのは割とニュートラルに使えるので、いろいろなところで実際に使われています。

ダブリンコアについての日本語でのきちんとした説明がありません。私がさぼっているからといってよく怒られるのですけれども、意外とありません。日本国内にダブリンコアのことをきちんと支えるための組織がないことも、日本語による情報が不足する一つの理由です。今の数分の説明なのですけれども、ダブリンコアのエッセンス、あるいはダブリンコアの心というのはこれだけで充分であろうと思います。

というところで、時間が少し過ぎてしまいましたので閉じたいと思います。どうもありがとうございました。