

# 電子化と情報化の概略

参考：日本の考古学リソースのデジタル化 <http://www.amy.hi-ho.ne.jp/mizuy/>

## 道具からメディアへ

コンピュータは「道具」だと思われているようです。しかし情報化や高度情報化という観点から見ると、コンピュータは「メディア(媒体)」です。コンピュータや電子情報技術は、何らかの著作物作成のための道具であり、見方を変えると著作物を流通するメディアでもあります。「道具」の面だけ捉えていると、情報の公開や共有といった発想は生まれてこないかもしれません。発想の転換が必要です。コンピュータ活用の様態は、以下のような発展段階を辿ると考えられます。

情報の処理 作成 整理 検索 公開 共有 知識の輪廻

コンピュータには、ハードウェアとソフトウェアという二面性も知られています。コンピュータ本体はハードウェアに違いありませんが、フロッピーディスクはどちらでしょうか。ディスクには、データが記録されています。データこそ、本質です。

情報化とは、文書(著作物)がコンピュータのデータになっていくことです。人文の場合、印刷物からデータに移行していくわけです。データはディスクからディスクへ、あるいはメモリへ、複製という輪廻を繰り返していきます。ディスクやメモリはデータの乗り物にすぎません。データの永続性こそ重要です。

情報がデータとなり、実体から遊離していくような感覚は、バーチャル化ともいわれています。バーチャルリアリティは仮想現実と訳されていますが、バーチャルとは、事実上現実と等しい、という意味です。決して仮の姿という程軽い意味ではありません。誰も、預金通帳の数字を見て、それが現金と等しいことを疑わないはずで、それは立派なバーチャル化です。数字と現金の関係を理解するのは、現代人の条件です。

今後、報告書がデータになっていくということは、ある意味で報告書がバーチャルな存在になっていくということです。この場合、バーチャル化は、あまり人気がありません。どんなに、技術的に問題がないとか、当分の間、紙の報告書を無くすわけではないと説明しても、多くの場合聞き入れてもらえません。

たとえ話を続けると、通貨を通貨たらしめるのは、信用です。信用は、裏付けによって得られます。昔は、金(ゴールド)が裏付けでした

今、フロッピーといえば、3.5インチで1.4MBの磁気ディスクのこと。ちなみにフロッピーの耐久性は数年程度。1.4MBといえば、単純計算すると約70万字に相当。

MO(エムオー)とは、3.5インチ光磁気ディスクのことで、230MB、640MB、1.3GBなどの種別があります。MOの耐久性は50~60年以上。

CD-ROM(シーディーロム)は、物理的、形状的、製造工程的に、音楽CDと同じもの。12cm径で、容量は650MB。ポリカーボネイトで出来ているので、紙より丈夫。但し、レーベル面はデリケート。耐久性は100年以上。

CD-Rは、記録メディアとして開発されたCD。形状的にも電子的にもCD-ROMと同じもの。書換えや消去が出来ないので、公文書にも適している。耐久性はCDと同じ。

ハードウェアは装置ないし機器、ソフトウェアはプログラム及びデータの総称。システムとしてのコンピュータから、作業に応じて換える部分を分離したものが、ソフトウェア。

バーチャルリアリティとは、映像や音や振動によって、現実世界を再現するものです。

日本国政府にも「電子政府」「バーチャルエージェンシー」という構想があります。これは、現実の施策です。

が、金本位制が停止された現在では、国家が裏付けです(国家の経済システムが順調に運営されている状態が条件です)。今でも金は国家の金庫に積み立てられていますが、世界経済に占める金の比重は低下しています。

電子的な報告書にも、マネーの世界とのアナロジーが成り立ちそうです。紙本位制と金本位制のアナロジーです。今後、紙の媒体は、バーチャルな媒体の担保になるでしょう。紙媒体は担保として書庫に積み立てますが、実際に流通するのは電子媒体になるでしょう(ある意味で、担保と交換可能な存在が、PDF - 後述 - だと思います)。通貨の世界で起ったプロセスが、文献の世界でも起るわけです。

デジタルな媒体に不可欠なのは、信頼できる技術と装置ですが、それらを維持する社会や習慣も重要です。デジタル情報が技術的に信頼に価するものとなっても、それが社会的に受容され、継承されていかないと、意味が無いわけです。通貨における国家の存在と同じように、文明社会が安定して維持・発展している状態が、文献電子化の裏付けとなります。

## ハイパーテキスト

メディア路線の原点は、大量の知識(文書)を自在にナビゲートし、迅速に閲覧したいという願望から来ています。最初、それはマイクロフィルムのアイデアでした。任意の頁を、思うままに素早く探し出して表示したいという願望は、コンピュータが利用できるようになってから後に、ハイパーテキストの概念にまとめられました。簡単にいうと、(デジタル化された)文書中の任意の語句から、別の箇所ないし別の文書へジャンプし、表示を切り替えるという仕組みです。これによって、文書は文書固有の一連の文章の流れから解放され、複数の文書間を自在に行き来することができるようになりました。

端的にいうと、引用や参考文献があれば、その場で原本が読めるようにするという事です。また図や写真を示す時、原本を借りたり、コピーして貼込むのではなく、オリジナルのデータにリンクを張り、新たな文書に埋込むこともできます。そうなると、実データのやりとりの必要は無くなります。こうして文献間の電子的なリンクを拡張していけば、世の中の総ての文献が、一続きの書物(全集)のようになっていきます。

ハイパーテキストは、長い間、憧憬でした。囲い込まれた環境でそれを実現することはすぐ出来ましたが、それがインターネットとHTMLの出現によって、突然大衆のものとなったわけです。ブレイクしたのは1990年代中葉のことでした。今は、HTMLでは不十分だった機能を、XML化によって補っていきこう、という段階です。電子化・情報化とは、世界大のハイパーテキスト願望の実現の道程だといえます。

余談ですが、世界中で国庫の金(ゴールド)は、徐々に市場に放出されています。そのため、金の価格は長期低落傾向にあります。

PDF(Portable Document Format)は、印刷物と同一性を持った電子文書規格です。

紙の媒体でも、超長期保存は不可能です。もし千年報告書を作りたいなら、和紙に墨で印刷したらいかがでしょうか。現実には、複製を繰り返す(昔なら写本です)以外に、文化を継承する方法はないのです。

本気で「保存」を考えるなら、デジタル以外にありえませんが。音楽でも、映画でも、写真でも、既に実行されつつあることです。

フィルム利用のアイデアは、1930年代にラピッドセレクトアとして開発されていました。いわば、迅速情報選択表示機です。

1945年に発表されたMEMEXの概念は、後の電子文書時代を予見していました。

ハイパーテキストの持つジャンプの仕組みは、リンク、ないしハイパーリンクと呼ばれています。

インターネットは、世界中とつながったネットワーク環境を指します。いわゆる「ホームページ」は、ワールドワイドウェブ(WWW)が正式名称です。略してウェブともいいます。一連のウェブのトップページを、ホームページと言います。WWW文書の規格が、HTML(HyperText Markup Language)です。

## ユニバーサルアクセス

報告書電子化の動機を一言でいうと、情報へのユニバーサルアクセスです。ユニバーサルとは「普遍的」ということですが、具体的には「いつでも、どこでも、誰でも、コネが無くて、資金力に関係無く、そして情報のどの部分でも、必要なだけ、あまねく、公平に、利用しやすく」ということです。電子化は、ユニバーサルアクセス実現の決め手となるはずです。

ちなみに電子化の最終目標は、あくまで知的生産の拡大を図ることであり、知識の輪廻のプロセスを電子的なものに置換えていくことです。

### ユニバーサルアクセス

- 1) 資金・時間・距離・書庫といった制約のない情報入手手段を広く提供するユニバーサルサービス
- 2) 従来の報告書では収録できなかった広範な情報を収録して、情報のワイド化とリッチ化を図る
- 3) 情報の有効利用率を高めるため、情報を整理して分りやすく提供する情報デザイン、及びネットワークの活用
- 4) 検索性を高めるため、メタデータを整備する
- 5) 文献におけるユニバーサルデザインの実現

### 1) 文献入手と図書収蔵・検索にかかるコストの削減

古書店や図書交換会(?)には、立派な価格で報告書が並んでいます。都道府県レベルの機関やコネのある研究者は、寄贈を受けているから関係ないかもしれませんが、区市町村レベルや個人では金がかかってしょうがありません。たとえ資料を入手したとしても、手持ちの本が多くなり過ぎると、多すぎる本は探すのが面倒なので、時として無いのと一緒にだったりします。ちゃんと図書整理しておこうとすると、かなりエネルギーや時間を消費するはず。書棚の容量も限界があります。新築する時に書庫を作るとしても、金のかかる話です。

無論、本や資料が増えていく事情は分かっています。気がついた時に入手しておかないと、後で入手できるとは限りませんし、必要な資料本体が入手できなければ(入手しにくければ)躊躇なくコピーします。しかし、その循環を繰返していく結果が、溢れかえる書棚です。電子化は、こうした文献に関わる物質的な困難を解消してくれるはず。必ずです。

出版側から見ても、電子出版のコストは安くなります。絶版ということも原理的に無くなります。必要な時にデータを取り出して、閲覧すればいいわけです。こうして、今まで紙資料が占めていた空間(書棚)も、もっと別の、人間的な目的と余裕のために使うことができるはず。必ずです。

ユニバーサルアクセスは造語です。

ユニバーサルアクセスの概念は、近代図書館の思想に似ています。

電子化に対応できない人にとってはユニバーサルではないかもしれませんが。しかし従来型の方式の限界も明らかだと思えます。

知的生産の拡大を図るシステムとして、「ナレッジマネジメント」という概念もあります。主に企業の社内文書の共有システムを指しています。ナレッジマネジメントを、考古の世界で、オープンに実現できたら最高です。

メタとは、～を超えた、高次の、より高いレベルの、といった意味です。ここでは目録・要旨等の電子化とリッチ化を指しています。

ユニバーサルデザインは、福祉用語に端を発していますが、工業デザインの一つの潮流です。

報告書の実需はどのくらいあるのでしょうか。つまり研究者が、自分の研究に必要な資料です。しかも、必要な部分は、報告書のごく一部だったりします。需要に忠実に対応すること、つまりオンデマンドは、これからの生産の潮流です。

情報の所在を明らかにし、いつでも入手可能であることを保証しておけば(電子的手段によって対価はゼロに等しく...) やみくもに、使うかどうか分からない資料を収集する必要はなくなります。

電子化によるコスト削減は、当座のことばかりでなく、紙資源の有効活用、印刷・輸送・廃棄にかかるコストを長期的に削減し、環境コストを下げることに繋がります。地球に優しいメディアというわけです。

## 2) 情報のワイド化・リッチ化

電子化によって、今までは収録が不可能であったような幅広い情報(ワイド化)、豊かな情報が収録できます(リッチ化)。文章でも表でも写真でも、電子版においては、収録量の制約は(事実上)ありません。

特にカラー写真の情報量は、確かに電子化の威力の一つです。白黒写真とは、情報量が全く違います。また実測図や計測表には、時としてミスがあります。それを後で検証可能にするために、スケール入りの鮮明なカラー写真を収録しておけばいいのです。カラー写真を収録するだけなら、大した手間はかかりません。もし余裕があれば、3D(スリーディー)表現やGIS(ジーアイエス)もいいでしょう。他にもワイド化・リッチ化の可能性はいくらでもあるはずですよ。

## 3) 情報デザインとコミュニケーション

大量の情報を相手にする場合、それらを適確にナビゲートし、情報が有効に、かつ(なるべく)正確に伝わる技術が重要です。これは情報デザインの問題でもあるし、文章力やコミュニケーション術、あるいは情報の所在に関するデータベースの問題かもしれません(メタデータ)。

発掘情報や埋文情報は、特に日本では大量に存在しています。でも現状では、情報収集の効率が悪すぎます。情報デザインの質を高めていくことで、情報のロスを減らし、情報の伝達効率が高まることを期待します。情報が収蔵庫や報告書のどこかに眠ることなく、しかるべく活用されるためには、情報の伝え方を工夫する必要があるはずですよ。

後述するメタデータや、情報デザインでカバーできない部分は、電子的ネットワークによって、コミュニケーションの密度・効率を高めて対応していきます。具体的にいうと、WWWや電子メールの活用です。

## 4) メタデータの整備

メタデータとは、情報(報告書や遺跡)の所在や内容に関する、インデックス的なデータです。ここでは、書誌(目録)~要旨~抄録~文献本体、という文献サービスの階梯を意味します(文献本体とは、文献全頁の電子版)。報告書の電子化は、単体で存在するのではなく、こうした総合的な情報網の階梯の一部に位置付けられるべきです。

遺跡地図も同様です。遺跡情報も、浅く広いデータから深く狭いデータへの階梯が必要です。このアイデアを、ISEKIDASとして発表したこ

余談ですが、中国が本格的に近代化すると、知識の媒体としての紙と、トイレトペーパーの使用量が、生産の限度を越えて必要になるそうです。電子化への期待は、環境問題にも直結しているのです。

3Dは立体表現ということですよ。遺物を、正面の写真だけでなく、横も裏側も任意の位置で見れたらいいでしょう。正確な立体CGが作れるなら、なお最高ですよ。

GISは地理情報システムと訳されています。地図(平面図)上に、何らかの分布をグラフ化して重ねた表現ですよ。本格的にいうと、空間表現の可能なデータベースですよ。

インフォメーションマッピングというものがあります。主に文章の構成要素を意識し、明確な構造化を図り、それを項目やラベルの形で明確化することを意味します。メッセージが、なるべく歪み無く、正確に、迅速に伝わるのが目標ですよ。

電子メール(Eメール)は、組織の情報化の起爆剤になります。

ここでいう抄録とは、(字義通り)文献を部分的に抜き出したものを指します。

ISEKIDAS(Inventory of Sites by Excavation and Kinds of Implements ; Documents for Archaeology Service) 要するに遺跡地図・台帳の拡大版ですよ。

とがあります(『報告書の電子化』1997、考古学ジャーナル)

### 5) ユニバーサルデザイン

いわゆるバリアフリーは、身体能力の如何に関わらず、誰でも生活しやすい環境を整備していくことですが、もう少し普遍的に、誰に対しても有効で使いやすいデザインということで、より包括的なユニバーサルデザインが注目されています。

電子化は、情報のユニバーサルデザインの実現にも有効なはずで、デジタル技術は、身体能力の拡張を可能にする技術でもあります。記憶力も拡大しますが、例えばディスプレイによる拡大表示は、視力の弱さをカバーします。テキストデータは、音声での読書も可能にします。

誰にでも分りやすい「情報デザイン」も、ユニバーサルデザインに通じますし、さらに「マルチメディア」も、分りやすさという同様の動機を持っています。

マルチメディアは、元来は、映像・音声・テキスト等を自在に組み合わせた表現ということです。視聴覚教育の発展形としても、教育・学習効果が高いということで注目された経緯があります。但し通常の「マルチメディア」は、知識の拡大というよりも、底辺の拡大、即ち大衆的な市場の拡大に寄与するものです(テレビ番組のように)、制作コストもかかりますし、我々としては「マルチメディア」よりも、まずドキュメントの電子化を実現するべきです。

### 写真のデジタル化

電子化とは、言ってみればバーチャル化ですから、その信頼性と共に情報再現性が問題とされるでしょう。目標は、実物にあたることなく、どこまで実物の観察に匹敵する情報を取得できるか、でしょう。照明の具合が関係してきますが、肉眼で見る程度の情報量を写真でカバーすることは技術的には可能です。コンピュータでは、拡大が自在ですから、ルーペで覗くくらいの情報量は容易に表示できます。

写真を贅沢に利用していきまると、大量の撮影にかかるフィルム代・現像代がかさんできます。コストの見地からいっても、デジタルカメラを活用すべきかもしれません。一方、フィルム写真のデジタル化には、フォトCDが最適です。コストは一コマあたり150～170円くらいです。

なお、写真の保存、特にスライドの退色を防ぐためにも、デジタル化、フォトCD化は有効です。フィルムの長期保存には、本格的な低温設備が必要になります。そうした設備投資が一般的に可能だとは、とても思えません。フォトCDなら一般化できます。デジタル化しておけば、情報は変化しませんし、フォトCDはCD-Rですから、通常の保存環境で

代表的な例でいうと、車椅子でも生活しやすい住宅、公共施設の整備ということですが、無論、老化や他の身体的な問題にも対応しますし、小物から都市計画まで、幅広く対応する概念です。

ディスプレイの性能は、将来は紙に印刷したものと同等に、高品位で自然なものになるでしょう。

デザイン重視で、字のサイズを小さくする傾向がありますが、止めて欲しいです。PDFなら拡大表示できますが...

マルチメディアをあえて訳せば「多媒体」でしょう。視覚的に見せることで、情報伝達の効果を高めるデザインです。

マルチメディア風デザインの特徴は、色使いです。やや暗い背景が好まれています。

デジタルカメラの中心は、300万画素オーバーに移行しつつあります。ここまで来ると、A4判半頁大の扱いも現実的になります。次に買うカメラは、もう決まりでしょう。

なおデジタルカメラのデータは、(加工は度外視して)早い内にCD-Rに焼いておくことをお勧めします。確実な保存のためです。

フォトCDは現像所経由で発注できます。1枚のCDに100コマ入ります。

常温では、適湿を保ってもフィルムの退色は避けられません。本当の保存を考えると、常温では無意味です。

も少なくとも100年の耐久性があるとされています。100年以内には、複製を作り直す必要があるようですが、デジタルですからコピーも簡単です。ちなみに通常の写真でも、超長期保存を考えると、複製を繰り返す以外に方法はありません(コダックのマニュアルに書いてあります)。

フォトCD化は、急務です。これは埋蔵文化財の保護保存の一環です。毎年、フォトCD化を計画的に事業として進めたいのです。当然、何が写っているかのデータベースも用意していく必要があります。

## 報告書の電子化

報告書を電子化していくとすれば、基本要件が4点挙げられます。

1. 報告書のレイアウトを再現するフォーマット ..... PDF
2. 電子時代のハイパーウェア ..... HTML(文章主体)
3. 互換性の高い生データ ..... Excel 4.0、JPEG 等
4. インデックスデータ .....(いわゆる報告書抄録 + )

ここで中心になる規格は、PDF (Portable Document Format) です。PDFは、印刷物のレイアウトを完全に再現する電子文書規格です。拡大縮小やハイパーリンクといった電子文書的な機能も完備しています。PDF自体の制作コストは殆どゼロです(HTML版で画像を含めると、大変な二度手間になってしまいます)。PDFは、印刷用のDTPデータから自動的に変換することで作成します。言い換えると、印刷用のデータ制作過程が、そのまま電子文書の制作過程になってしまうわけです。手軽というか、手品みたいな話です。PDFは、印刷データのバックアップともなります。PDFは、文書「表示」の確実な共有システムです。こうしたPDFの性質は、紙文書偏重からの脱却の決め手になります。

PDFの制作は、印刷会社に、印刷発注の時に同時に頼んでしまえばよいのです。自前でDTPを実践している現場であれば、自前でPDF化してもいいのですが、完全なDTPの実践は一般には不可能に近いことですし、校正の関係で、PDF化のタイミングは下版時、即ち印刷直前のデータから作る必要があります。その時点では、校了データは印刷会社の手元にある可能性が高いわけです。印刷会社にPDF化をやってもらう作業について、余分なコストは殆ど発生しないはずですが、ハイパーリンクを作り込む場合には、多少のコストがかかるでしょう。

PDFの仕様例を示します。解像度別に、商用印刷対応レベル(高解像度版)、ローカルプリンタ対応レベル(中解像度版 - 本来の閲覧用)、さらに中解像度版の圧縮率を高めた、オンライン版の3種を制作します。オンライン版の解像度は、もっと下げた方がいいかもしれません。

建設土木の分野では、デジタル写真が提出文書に利用され、データベース化が進んでいます。建設省の指導に基づくものです。

フォトCDの利用に際しては、基本的な仕様を理解している必要があります。

DTP(DeskTop Publishing)は、ポストスクリプトベースの編集システムです。完成したDTPデータは、そのまま印刷工程に回すことができます。

DTPは電子化への道程かも知れませんが、文書の100%電子化を学習する機会です。

PDFの閲覧には、無償配付のAcrobat Readerが必要です。最近では、コンピュータ標準装備になりつつあります。最新版は4.0です。

PDFの作成には、製品版のAcrobat(約25000円)が必要です。

実はワープロからPDF化することも可能です。でも図版ページが多いと、あまり現実的ではありません。また文字組みの完成度も、ワープロで実現するのは困難です。やはり出版物ならDTPです。

PDFには、しおりブックマークという仕組みがありますので、目次をこれに反映させます。それ以上のハイパーリンクの作り込みは、余程余裕がある場合に限られます。

Acrobat Distillerでも、プリプレス最適化、プリント最適化、スクリーン最適化という3段階を用意しています。

高解像度版の解像度は、印刷用に用意したデータそのままです。

	高解像度版	中解像度版	オンライン版
写真	300 ~ 350dpi	150dpi	100 ~ 150dpi
JPEG 圧縮	最高画質	標準圧縮	低画質
2 値画像	800 ~ 1200dpi	600dpi	300dpi
互換性	任意	3.0 互換	3.0 互換

互換性とは、Acrobat 3.0 互換ということです。現在は Acrobat 4.0 が登場していますが、旧バージョンとの違いは、日本語フォントの埋込みと、写真のプロファイルの埋込みです。基本的な閲覧版は 3.0 互換データにしておくべきでしょう。しかし閲覧環境としては、4.0の方が快適です(2 値画像の表示が滑らかです)。

### マークアップ型の電子化

電子文書のための規格としては、マークアップ言語 (Markup Language) が本流かもしれません。具体的にいうと、SGML、HTML、XML、XHTML です。この中で最も普及している HTML は、SGML に基づいて作られたウェブ用の規格です。XML は、SGML の規格を見直し、インターネット時代に適応して利用しやすく改良したものです。HTML も見直されて XML 準拠になり、XHTML が登場することになりました。ややこしいので、まとめて ML 系文書規格としておきます。

ML 系文書規格の (本来の主旨としての) 特徴は、

- 1) ファイルの全ての部分がテキストである...処理しやすく、可読性が (経時的にも) 高く、ハードウェアの規格を問わない。
- 2) 文章に論理構造が埋込まれ、物理的書式 (スタイル) が分離されているから、データ構造がすっきりして処理しやすい。
- 3) いかなる表現形式にも対応できる (開発次第であるが)、画像等は、外部ファイルへのリンクという形でいくらでも対応できる。つまり汎用性と拡張性が高い。

ML 系文書は、文書データの再利用や高度利用を目指して開発されました。最大の特徴は、文書構造の明確化と書式の分離です。異なる書式を適用することで、異なるレイアウトにフォーマットすることができます。但し物理的な「表示」を確実に共有する目的には、向きません。

HTML の登場で衝撃的だったのは、ハイパーリンクが、簡単な仕組みで実現されていたことです。HTML は、最初に実用化され、最初に大衆化したハイパーテキスト環境となったわけです。

日本語フォントの埋込みとは、表示・印刷可能なフォントデータを、PDF 内に埋込んでしまうことです。これで日本語フォントが用意していないコンピュータでも、表示できるわけです。

写真のプロファイルとは、正確な色再現のための仕組みです。

滑らかな表示は、PDF の特徴の一つです。

マークアップといえば他にも TeX (テック、あるいはテフ) という規格がありますが、DTP 用の規格です。

全ての HTML が XML 準拠になるわけではありません。

SGML - HTML - XML の関係は分りづいかもしれませんが、SGML / XML は、電子文書フォーマットを定義するための、マークアップ言語の仕組みを定めているものです。HTML は、SGML によって定義された、具体的な文書フォーマットです。文書フォーマットの定義 (文書型定義) を DTD といいます。HTML には、HTML の DTD があります。

ML 系で、文書レイアウトを完全再現する可能性は、ないわけではありませんが、技術のあり方としては、少々期待薄です。但し XML で DTP という話なら、可能かもしれません。