

デジタル書誌学の諸課題

Some Thoughts on "Digital Bibliography"

久須美雅昭

はじめに

①問題の所在

②書誌学の定義

③デジタル資料の定義

④デジタル・メディアをめぐる

⑤映像書誌学

⑥データベースをめぐる

おわりに

【論文要旨】

研究資料のデジタル化は、とりわけ人文系研究の領域において、新たな方法論の開拓や、これまでにない共同研究スタイルの確立などにつながる大きな可能性を秘めている。その反面、デジタル資料への依存は原本の喪失という危険な側面も併せ持っている。デジタルであることはすなわち、デジタルを解読するハードウェア、ソフトウェアに依存することであり、このようなハード、ソフトは市場の原理で極めて移ろい易いものだからである。

まさに資料デジタル化を始めた世代の責任として、1000年先の世代に研究資料を引き継ぐという観点からデジタル資料の特質を理解し、その妥当な扱い方を検討する必要がある。

それには文書資料における書誌学を下敷きにすることが有効であろう。ここでは、仮にそれをデジタル書誌学と名づけ、デジタル資料の多面的な定義、あるいはハード、ソフトへの依存性を考察する。

また、動画映像資料のデジタル化に先立つべき、映像書誌学の必要性にもふれる。

最後に、デジタル化の実作業の場としてのデータベースについてトヨタ財団の例に基づき考察し、デジタル資料の扱いにかかわる様々なヒューマン・エラーの可能性を明らかにする。また、データベースの延長上にWEBアーカイビングと呼ばれる最近のデジタル資料生成の方向性も検討する。

付録として、筆者が開発した画像データベース構築のためのツールについて紹介する。

はじめに

本稿は、2000年9月29日に歴博の「民俗学的画像に関する基礎的研究」のゲスト・スピーカーとして行った「デジタル書誌学と画像DBツール」と題する口頭発表をもとに、その後の状況変化や自分自身の考え方の展開を踏まえて執筆した。

歴博の研究会で行った口頭発表の趣旨は、筆者自身が十数年前に着想を得て、白石昌二郎氏（情報環境デザイン(株)）と共同で開発した画像データベース構築のためのツール（Album Tools）を紹介するとともに、この10年間に問題として意識されてきたデジタル化資料の原本をめぐる諸課題についての問題提起を行うものであった。本稿では、ソフトとしての完成度が未だ十分とはいえない画像DBツールの紹介は付録とするにとどめ、問題提起の方に焦点を絞ることにしたい。

最初にお断りしておくが、筆者は学術研究の訓練を受けた専門研究者ではない。トヨタ財団という民間助成財団で、プログラム・オフィサーとして、長年、研究助成業務を担当してきた者である。この過程で、理系文系を問わず多くの研究に接し、学術論文には描かれることのない、研究者のデータ収集作業などの実態を数多く垣間見てきた。また1985年から90年にかけて助成財団センターに出向し、組織の立ち上げと民間財団のデータベース構築などの実務に携わった。この時代はちょうどパソコンの興隆期であり、筆者自身、パソコンの進歩に合わせて、数種のデータベース・ソフトでプログラム開発を行った経験を持つ。

このような経験の中で、近年急速に進められつつある研究資料のデジタル化をめぐって問題を感じることが多くなってきた。それは、とくに文科系研究者に目立つ傾向なのだが、デジタル化に対する過大な期待がしばしば見られるということである。要は「デジタル・データはアナログ・データとちがって複写を繰り返しても劣化しない」という特性が拡大解釈されて、「デジタル・データは不滅である」という間違った認識が広まっているということなのだが、これは100年先に禍根を残す危険な傾向と思われる。すなわち、われわれの時代に、先の世代に引き継ぐべき研究資料の「原本」というものをことごとく失ってしまうのではないかと危惧されるのである。そこで、本稿では、筆者がなぜこのような危惧をいだいているかということを出発点として、そもそもわれわれはデジタル化資料というものをどのようにとらえるべきなのかということ、書誌学の比喩を適用しながら論じてみたい。

なお、表題にいうデジタル書誌学⁽¹⁾というのは筆者の造語である。しかし、最近インターネットの検索を通じて、伊藤政行氏が「デジタル書誌学」という用語を、書誌学的方法論——すなわち原典に依拠する——をデータベースやインターネットといったコンピュータ・メディア上で展開していく⁽²⁾という意味で使っていることを知った。筆者の意図はこれとは若干ずれるかも知れないので、「デジタル化された資料に対して文書書誌学の比喩にもとづき検討を加える学問」の略記としてデジタル書誌学を使うことを了解いただきたい。

①……………問題の所在

デジタル化の何が問題なのかを指摘する前に、コンピュータが人文系の研究に飛躍的な発展をもたらしていることはまず評価しておきたい。とりわけ重要な役割を果たしているのがデータベースとインターネットであり、さらにその背景となるマルチメディア化の進展である。

データベースについていえば、80年代後半くらいから個人ユーザーがワープロを使いこなす程度のコンピュータ・リテラシーで容易に本格的データベースが構築できるようになってきた。この結果、コンピュータがなければおよそ実現しなかったかも知れない研究成果がいくつも生まれてきている。ひとつの例としてトヨタ財団のWEBサイトから公開されている「日本の産業遺産ヴィジュアル・データベース」(<http://www.ied.co.jp/isan/sangyo-isan/JS-TOP.htm>)をあげることができる。ここに収録された約7000件の産業遺産のカルテ化に際してはR:Base、桐、Accessなどのパソコンで使えるデータベース・ソフトが威力を発揮した。ある意味では、産業遺産データベース・プロジェクト自体、パソコンの進歩と普及を背景にはじめて構想し得たものであった。

また、90年代に入ってからインターネットの普及はこのような膨大なデータを含むデータベースを、印刷媒体を経由せずに直接世界に配信する道を開いた。

さらにマルチメディア化は、これまで異質の研究資料とされた文書、画像、音像などをデジタルという共通の基盤の上で一元的に扱うことを可能にした。

人文系の資料研究などでは、とかく希少資料の発見者がその資料を抱え込んで、断片的論文を小出しにして業績をかせぐようなことがあったと聞く。しかしこれからは、初出資料はただちにインターネットを通じて広く共有し、様々な専門家がその資料をもとに集約的かつ累積的に成果を出していくような、データベース型共同研究が新しい研究スタイルとなるであろう⁽³⁾。

では、いったい何がデジタル化の問題なのか。それは第一に、ハード、ソフトを含めたデジタル・システムはそもそも資料の超長期保存に向いていないということ。第二に、書誌学的な観点に立った場合、資料原本の遍在もしくは不在により基準原本が確定できないということである。

まず、第一の点だが、テープ、ディスクなどのデジタル・メディアは一般の期待に反して、実は短寿命なのである。ホームビデオのテープが10年で劣化することは既に多くの人が経験として知っている。フロッピー・ディスクやハード・ディスクが物理的に壊れるとどうにもならないことも皆承知している。少なくとも磁性メディアは磁気に弱い。ここで、CD-ROMやDVDの寿命はもっと長いという反論があるかもしれない。確かに推定寿命100年以上ともいわれている⁽⁴⁾。しかし、メディアは残っても読取装置の方が先に消滅してしまうのである。80年代に使われた8インチや5インチのフロッピー・ディスクは、かりに今保管されていたとしてもそれを読み取る機械を探し出すことは極めて困難であろう。同様に、秒進分歩といわれるコンピュータの世界で100年先に現代のCD-ROMやDVDを読み取る機械が残っていると考えるのはまずナンセンスなのである。

さらに、デジタル・メディアはフィルムなどと違って、それ自体を見て内容を理解できるものではない。このことが超長期保存に向けて何世代にもわたって人間が関与するような仕組みを考えるとき、実は致命的な短所となる。

次に、第二の資料原本としての問題である。デジタル化された資料を歴史研究などの典拠として使う場合を考えてみよう。デジタル化の元資料が存在している場合はそれが典拠となるべきことは異論がないだろう。しかし、なんらかの理由で元資料が失われた場合、あるいはデジタル写真やプログラムのソースコードのように、そもそもがデジタルの一次資料である場合には、原本の特定は極めて困難となる。なぜならデジタル・データの特性として、複写によって同一の原本がインターネット上に遍在する事態も大いに起こり得るからである。ならばどれを原本としてもいいではないか、というところはいかない。もうひとつのデジタル・データの特性として改竄は⁽⁵⁾いとも容易だからである。

また、データベース・コンテンツにおいては理論的に書誌学的な意味での原本は存在しない。これは後に詳述する。

以上のように、資料のデジタル化は、学術研究の世界に革命的な飛躍をもたらす可能性をもつ反面、100年、1000年という歴史の中で取り返しのつかない失敗となる可能性もはらんだ両刃の剣なのである。

②……………書誌学の定義

デジタル化資料に書誌学の比喩を適用すると述べたので、まず書誌学の定義にふれておく。長澤規矩也氏の『古書のはなし——書誌学入門——』（1976年、富山房）によると、書誌学は図書館学と同義で英語の bibliography の訳語だという。この言葉は大正末期にできたもので、それ以前は、明治後期から、史学者によって、書史学とよばれていたというから実は意外に新しい。そこで bibliography をブリタニカ百科事典で調べてみると、図書の分類整理の学 (enumerative, systematic, or descriptive bibliography) とモノとしての図書の成り立ちを調べる学 (critical bibliography) との両義にまたがっている⁽⁶⁾。

長澤氏が書誌学の研究対象として整理した項目も、以下のようにモノとしての成り立ちから、分類整理、さらには流通にいたるまで広い範囲にまたがっている（前掲書、p. 153）。

- 一 図書の定義、範囲、種類、起源、発達。
- 二 図書の材料、形態（大小、様式）、装訂（釘は同音の誤用）、付属品。
- 三 書写及び印刷の材料、様式、方法、種類、歴史。
- 四 内容（的本）の成立（著述、編修、翻訳、図表）、種類、校訂、伝来、存亡。
- 五 図書の集収、保存、分散等に関する事情、方法、歴史。
- 六 文庫と図書館との相違、発達、種別、建築。
- 七 図書整理の原理、方法、歴史。
 - ア 選択法、イ 目録法、ウ 分類法
- 八 著作権・出版権・販売権等、図書に関する法律規則。
- 九 図書を対象とする各種の企業（編修、印刷、製本、出版、販売、貸本等）⁽⁷⁾

これらの項目からの筆者の理解では、書誌学とは、書かれた内容を云々する学問ではなく、それに先立ち、書かれた媒体の成り立ちを明らかにすることで、内容の信憑性を裏付ける基礎的な学問

といえるのではないかと思う。そして、その比喻として、デジタル書誌学は、コンテンツのクリティックには踏み入らず、コンテンツを運ぶ／担うデジタル・メディアに注目し、そのメディアの成立過程、技術的背景、などを検討する基礎的な取り組みであると位置付けたい。

長澤氏が掲げられた項目は、「図書」を「デジタル資料」と置きかえるだけでかなりの部分が先に述べたデジタル化にかかわる今日的課題に適用できると思われる。しかし、本稿の限られた紙数では逐条的にすべてを論ずることは無理なので、以下ではいくつかの項目に絞って検討することにした。

③……………デジタル資料の定義

デジタル・データとはつまるところメディア上に記録されたオン・オフの2値による2進法の数列である。その中には、点の集合に還元された画像データもあり、数字コードに置きかえられた文字データも含まれる。ここでいうデジタル資料とは、「デジタル化された資料」の略記であるが、その態様として、①実物のデジタル写像、②コンピュータ生成像、③文字データ、④プログラムそれ自体という4つが考えられる。以下、それぞれの態様について検討し、その総体としてデジタル資料を定義することにした。

(1)実物のデジタル写像

学術研究の観点から見るとデジタルの大きな効用のひとつはマルチメディアである。すなわち、文字、画像、音像を全く同じメディアに記録し、操作し（つまり分析、総合し）、パソコンなどを通して表現することが可能となったわけで、これは学術研究の方法論という点ではまさに革命的なことであろう。⁽⁸⁾

デジタル資料という場合、物理的なメモリー空間を最も消費するという意味で、現実世界を写し取った画像（静止画、動画）、音像などのデジタル化写像としての様態が第一にあげられる。

1 静止画像

デジタル化静止画像とは、スキャナー、デジタルカメラなどをとおして、既に存在している絵画、写真、文書等をデジタル化したデータであるが、最近では建築物や景観などを直接デジタルカメラで撮影した、すなわち一度も印刷されたことのないデータも多くを占めるようになっている。

そこで、元画像の所在、状態等がデジタル化資料原本としての観点から重要となろう。少なくともそのデータが何かを転写した二次資料なのか、最初からデジタルとして生成された一次資料なのかを識別するルールが必要となろう。

また、解像度および画像圧縮の有無やそのフォーマット、あるいはカラー問題などの技術的問題がデジタル書誌学の記述対象となろう。⁽⁹⁾

2 動画像

動画像は簡単にいえば1秒あたり数十枚（テレビでは30枚、映画では24枚）の静止画像の連続であるが、デジタル化に際しての圧縮技術は静止画のそれとは異なる。静止画の場合と同様に、原本の所在や圧縮技術などがデジタル書誌学の記述対象となろう。原本に着目すると8ミリ、16ミリ、⁽¹⁰⁾

35ミリなどのフィルムからデジタル化したデータと、放送録画、撮影ビデオなどフィルム以外のソースからデジタル化したデータの2通りが考えられる。ビデオが、フィルムと決定的に違うのはその内容が肉眼では推測できないということ。このことからフィルムとビデオテープとは書誌学的にはまったく異質のメディアと考える必要がある。

また動画像の場合、これを学術研究の資料として位置付ける方法論的な検討そのものが今までになされてこなかった。理由は単純で、紙ベースの論文では「動き」そのものが記述できないからである。従って、動画像をめぐるデジタル書誌学より先に、映像書誌学というもうひとつ別の体系を考えねばならない。これについては節をあらためて述べる。

3 音 像

民族音楽の録音やインタビュー記録などは学術研究の一次資料とされる。動画像とちがってこれらが資料となりえたのは、採譜、翻字、発音記号など、完全とはいえないまでもある程度の記号化が可能であったからである。

録音メディアには蠟管、レコード盤、フィルムのサウンドトラック、テープなどがある。デジタル化された資料には、これらアナログ録音からの変換データと、最近のICレコーダなどによる直接のデジタル録音データとがある。ここでもやはり、原本の所在や圧縮技術などがデジタル書誌学の記述対象となろう。

(2) コンピュータ生成像

デジタル化資料として、今後もしかしたら実物のデジタル写像を量的にしのぐかもしれないのが、コンピュータが生成した、いわば架空の画像である。

1 数値から生成される像

架空の画像の筆頭はコンピュータ・グラフィックス（以下CG）である。その製作過程には図面から立ち上げていくモデリング、あるいは立体像のスキャニングによる数値化などがある。この場合の実体は、立体表面を微細な三角形に分解しその各頂点を表した三次元座標の膨大な数字のセットである。

2 数値以外による操作画像

実は、コンピュータが生成する画像は上記の狭い意味でのCGにとどまらず、はるかに裾野が広い。たとえば、デジタル化された実写像をコンピュータ上で合成した画像、あるいはモーフ画像といった異なる画像の中間像をコンピュータで生成しあたかも連続的に変化しているように見せるものなど様々なパリエーションが広義のCGとして考えられている。

また、リモートセンシングにおける赤外線画像、X線画像など可視領域以外での画像、さらにはそれがコンピュータにより補正された画像、探査衛星からの伝送画像の補正画像などがある。これらは虚構ではないという理由でCGとは呼ばれないと思う。しかし、古典的な意味での「写真」とはもはや別の範疇の資料としてとらえられるべきであろう。

これまでに述べた「実物のデジタル写像」と「コンピュータ生成像」とは、実際には截然とは分けられない。たとえば、CG画像も高精度にプリントされれば写真となるし、モーフィングも動画像として放送録画に変換されうる。さらにたとえば、実際の人体を磁気共鳴撮像（MRI）で輪切りに

し、これによって得られた膨大なデータをCGのように再構成した場合、得られた画像はいったいどう位置付けられるのだろうか。⁽¹¹⁾そもそも古典的な写真も所詮現実の影であり印画紙上の粒子の分布にすぎないという意味では、それも含めて画像資料とは何かという哲学的な命題にまでさかのぼることになるかもしれない。

ともあれここではデジタル書誌学上の課題がひとつあげられる。それはデジタル化の元資料と、デジタル化に伴うズレ、さらにはデジタル化以後の改変ということである。たとえば、東大のデジタルミュージアムの中では、デジタル化以後の改変は、「デジタル修復・復元」という名でポジティブにとらえられている。⁽¹²⁾もし、1000年前の原本がしだいに劣化し内容の判別もできないようになっていくとするなら、2001年の技術で復元したデジタル化資料が「デジタル原本」となり、1000年後に引用の典拠となることも大いにありうる。そこで重要なのは2001年の時点でデジタル原本のオーセンティシティをどう担保するかということであろう。たとえば、技術的には電子透かし⁽¹³⁾のような手法もあるが、むしろ運用制度的に、インターネット上で公開される画像資料にはすべて地紋を刻印し、コピーであることが明らかなようにした上で、原本は個別に配信するなどの方策が現実的ではないだろうか。

また、デジタル化に伴うズレについてはプラス・マイナス両面がある。たとえば、遺跡発掘現場では最近デジタルカメラが多用されるが、CCDの集光能力のおかげでしばしば肉眼で見る以上に鮮明な像が得られるという。マイナス面では、たとえば長澤氏は実物の写真化によって微細な情報が失われたケースを紹介している。⁽¹⁴⁾この場合はデジタル化以前の写真化の段階ですでに問題があるわけだが、デジタル化にも同様の問題がある。古文書の研究者に聞いた話では、実物では筆順を判定することができるがデジタル資料ではそれができず、しばしば文字の判読に苦勞するという。

さらに、純粋にコンピュータが生成した画像の場合、何をもって原本とみなすかが課題となろう。たとえば、特定のアプリケーションにより生成されたデータは、そのアプリケーションにより解釈されない限りは意味をもたない。その場合、データ+アプリケーションソフト+それが参照するライブラリー+OSまでの一連のセットがそろわないと画像は再現できない。このセットをもって原本とみなすのは現実的ではないから、結局、たとえばCGの結果をビットマップ画像に変換した、その1枚の画像ファイルが原本とみなされることになるであろう。

(3)文字データ

デジタル・データとしては、画像データのようにそれ自身は無意味な01の羅列に過ぎないものもある一方、いわゆるテキスト形式として意味のある文字列に変換できるものもある。こうしたデジタル化された文字列については書誌学上の様々な課題が予想される。その最たるものはいわゆる「文字化け」の問題で、これは80年代以降ワープロ（専用機、PC用含めて）が爆発的に普及した間に文字処理方式そのものが大きく変化してきたことに起因する。

元来コンピュータはアメリカが開発をリードしたものであり、データ交換の標準化のため1963年に設定されたASCII（American National Standard Code for Information Interchange）ではその名のとおりアメリカで使われる英文字、数字、記号が7ビット（2の7乗＝128）の番地に割り当てられた。その後、1973年ISO（International Organization for Standardization）でこれが追認されるかたち

でISO646として世界規格化され、1987年以降ISO8859の系譜として拡張されてきている。一方、日本語の処理には8ビット＝1バイトでは足りないため、2バイト（2の16乗）の番地に7000弱の漢字、ひらかな、カタカナ等を割り当てたJIS漢字コード（JIS C 6226）が1978年に開発された。その後、83年、90年、97年の改定を経て最近のJIS X 0208：1997の規格に至っている。この日本語規格化の過程で注目すべきことは、JIS C 6226を当時の非力なパソコンに実装するための工夫としてアスキー・マイクロソフト社によりシフトJISというコード体系がつくられ、この一私企業の工夫がデファクト・スタンダードとなり、ついに97年のJIS改定では公式に追認されたということである。さらに、世界的には80年代後半から世界各国言語を一元的に収納するコード体系が模索され、32ビットに21.5億文字を収納できるISO10646の規格と、16ビットを単位とするUnicodeが別々に構想された。ここでもUnicodeはアップル、IBM、マイクロソフトなどの私企業コンソーシアムによる、パソコンへの実装を主眼においた規格であることを強調しておきたい。93年にはUnicodeがISO10646の部分に包摂されるかたちで統一され、それぞれISO10646-1：1993とUnicode1.1⁽¹⁵⁾として発行された。

2001年現在のパソコンではUnicodeが主流となりつつある。さらにUnicodeの先にはISO10646による21.5億文字の文字空間が用意されているため、古代文字の復元などの研究用途も含めて文字コードの割り当てに困ることはないであろう。ということは、Unicode化以降のパソコンで作成された文書については文字化けの問題はおのずと解消されるということである。それはおそらく旧コード体系を実装したマシンが世の中から消えてなくなる、いまからわずか数年後のことであろう。しかし、80年代から現在にいたるまでは上に述べたとおり文字コードの体系そのものが大きく変わってきており、この間に作成された文書については、異機種間のコード割り当ての差やあるいはアプリケーションソフトの差に起因する「文字化け」は不可避免的に内在している。実は現在もなお一般に利用されている多くのソフトはUnicode未対応であり、文字化けの可能性をはらんだデジタル文書が続々生産されているのである。

ところで旧コードで作成された古デジタル文書においては、パソコンに内蔵されていない文字を表現するのにいろいろと工夫がされてきた。たとえば■を当てて後ろにカッコ書きで文字の要素を記載するとか、外字を作成しこれを参照するとか、あるいは外字を画像化して文字の位置に埋め込むとかである。初期のワープロのように文書清書機という性格が強かった頃は、印字さえできればよかったため、こうした工夫も意味があった。しかし、今後こうした古デジタル文書をアーカイブ化して活用していくことを考える場合には外字処理の有無を詳細にチェックしていくことが必須の作業となろう。

ただし、その際、外字部分を単純にUnicodeに置き換えてしまうことは問題であろう。なぜなら、百年後のデジタル書誌学的な観点からは、古典的外字処理法は古デジタル文書の成立年代を推定する重要な手掛かりとなるはずだからである。

(4) プログラム

デジタル・データのもうひとつの態様は、アプリケーション、ライブラリー、OS等のプログラムそれ自体である。プログラムにはソースコードと、コンパイルされ機械語となったプログラムの二

段階があり、通常プログラムといえば、開発を終えたソースコードをコンパイルしたものをさす。ソースコードは意味のある文字列からなるので、後世においてアルゴリズム進化史のような学問領域が成り立つとすれば、その基礎としてデジタル書誌学的にきわめて重要な資料となろう。しかし、ソースコードは概して残らない。そもそもコンピュータの2000年問題⁽¹⁶⁾があれほどの大騒ぎになったのも、銀行などの基幹システムで動くことが実証されたプログラムについてはソースコードが破棄されていたからなのである。コンパイルされたプログラムそれ自体はもはや解読不能でデジタル書誌学の対象とはなり得ないであろう。ただし、それが現存するマシンの上で生きて動いているとすれば、デジタル書誌学上最も重要な解読ツールとなるはずである。

さて、以上に見たようにデジタル資料を定義する場合、①実物のデジタル写像、②コンピュータ生成像、③文字データ、④プログラムそれ自体というように少なくとも4つの異なる態様が考えられた。この中で、プログラムそれ自体をデジタル資料とみなすことは文書の書誌学の比喻からは考えにくいかもしれない。しかし、デジタル・データはいかなるものであれプログラムなしでは意味を表現し得ない。しかもそのプログラムとは人間によって記述された論理であり時代とともに姿をかえる作品でもある。そのような観点から、プログラムも含めて、あるいはプログラムとデータとが不可分のものとしてデジタル資料をとらえることが必要であろう。

④……………デジタル・メディアをめぐって

長澤氏によれば図書の定義につづく書誌学の研究項目として、図書の材料などのハード、印刷の方法などのソフト、さらに図書の流通などがあげられていた。これにならって、以下ではデジタル・データを担うメディア（媒体）について、ハード、ソフト、マーケットの面から考えてみたい。

(1)メディアとデバイス（ハード面）

コンピュータの特性として記憶メディア（＝デジタル・メディア）とそれを読み取るデバイス（＝コンピュータ本体ならび、これに接続できる機器の総称）とは不可分の関係にある。

記憶メディアにはROM、カードメモリ、内臓ハードディスク、磁気テープ、MO、フロッピーディスク、CD-ROM、DVD等、様々なものがあり、それぞれにデータの書き込みや読み取りを行うデバイスに依存している。即ち、デバイスがなければDVDといえども花瓶敷ほどの役にしか立たない。

これら各メディアの特性、商品としての栄枯盛衰、メディアとデバイスの組み合わせにおける互換性などの問題を明らかにすることはデジタル書誌学の一つの課題となる。

問題の所在でも述べたように、メディアとデバイスの移り変わりは極めて急激で、80年代にデータベース化に取り組みはじめたプロジェクトでは、必ずといっていいほど、メディアもしくはデバイスそのものの乗り換えを行っている。それは一方では、メディアやデバイスの性能向上が著しいため、ユーザー自ら進んで——しばしばデータコンバートの時間と労力を払ってでも——乗り換えてきた面がある。しかし一方、デバイスの寿命が家電製品なみ（ハードディスクは家電以下という⁽¹⁷⁾）に短いため、不本意ながら乗り換えを強いられてきた面も少なからずある。デジタル化とは、いわ

ば寿命10年程度のデバイスを永遠に乗り換え続けることともいえる。

(2) OS・ドライバ・アプリケーション (ソフト面)

メディアとデバイスの間には常にOSやドライバなどのソフトウェアが介在している。

OSにはUnix, Linux, Mac OS, Windows など様々あり、さらにそれぞれにおけるバージョンの違いも無視できない。というよりバージョンが違えば、たとえばWindows3.1とWindows95のように実質的にまったく別のOSと見るべきであろう。また、外部装置の制御にはデバイス・ドライバと呼ばれるソフトがあり、これにもバージョンの違いがある。

OSとドライバとの組み合わせには、俗に「相性」といわれる問題があり、機械のつなぎ変え、ソフトのバージョンアップは必ずといっていいほど障害を伴う。この相性の問題は必ずしも新しいものの方がいいとは限らないところが難儀である。

この上にさらにワープロやデータベースなどのアプリケーションソフトが乗り、それらのバージョンの違いが加わって、複雑さの度合いは人知を超えるところとなる。⁽¹⁸⁾

デジタル・データは実はこうした複雑なソフト群の組み合わせの上で生成、加工、保持されているのである。そのようなデジタル・データの生い立ちをいちいち記述することは不可能にせよ、少なくともこうした背景事情を明らかにしていくこともデジタル書誌学の一つの課題となろう。

つまらない例だが、たとえば、Macのワープロで作文し、メールを経由してWindowsのワープロで開き、それをデータベースにカット・アンド・ペーストで貼り付けたデータがあったとしよう。ここでは確実に丸数字やローマ数字の文字化けが起こっているほか、英文のアポストロフィやクォーテーションも半角カタカナに化けている可能性がある。こうしたさまざまな可能性を把握しておくことが大事だということである。

(3) 市場メカニズム

ユーザーにとってコンピュータのハード、ソフトの急速な変化はありがたいとばかりもいえない。⁽¹⁹⁾特に、企業ユーザーなどではLANに繋がるパソコンを買い換えるたびにOSやプレインストールソフトの足並みが乱れるのは頭痛の種である。にもかかわらず変化しつづけるのはなぜか。すべては資本市場の原理による。

CPUの性能はムーアの法則として知られているように18カ月で2倍になる。ハードディスクも同様に大容量化する。しかも安くなる。当然、マシンは瞬く間に陳腐化し、メーカー間で果てしなき競争が演じられる。その間、ソフトもマシンの性能に合わせてどんどん巨大化し、さらなるマシン性能の向上を要求する。ユーザーは、本当はソフトの必要以上の高機能化は望んでいない面もある。しかし、ソフトメーカーはそれでは商品が売れないので、マシンそのものにあらかじめ最新バージョンをプレインストールし、無理やりユーザーにソフトの買い替えを強いる。かくしてユーザーは愛用——すくなくともどこにバグがあるかわかっているという意味で——のソフトを使いつづけることすら困難な状況ができていたのである。

メーカーでは「互換性」ないし「標準」を標榜し、ユーザーが築いてきたデータ資産は継承されるという。しかし、必ず例外はあり、デジタル書誌学的にはこの例外の存在こそが問題になるので

ある。すでに、文字データの節でも述べたとおり、「標準」それ自体が多くはメーカーの都合から生まれたものであり、今日標準とされるものが100年の時を経てなお標準であるという保証はどこにもない。

このような観点から、メディアとデバイス、OS等ソフトの総体が市場の中で時間とともにどう推移していくかを詳細に記録することがデジタル書誌学の課題となるのである。

⑤……………映像書誌学

2002年2月現在、映像書誌学という用語はかなり強力なサーチエンジン Google によるインターネット検索でも見つからない。筆者と映像作家の孝寿聡氏とが1980年代前半頃からいっしょに考え出し使い始めたことばである。定義は、デジタル書誌学の場合と同じく、「動画映像資料に対して文書書誌学の比喩にもとづき、その内容ではなく資料としての成立過程、技術的背景などを検討する基礎的な試み」ということである。

この課題自体とても大きく、映像製作経験のない筆者には手に余るもので、また本稿の論旨からは多少ずれる。しかし、すでに述べたようにデジタル書誌学の中で動画映像資料の位置付けは不可欠であるため、最小限の要点に触れておきたい。

動画映像はリュミエール兄弟のシネマトグラフィカから起算して100年以上の歴史をもち、世界的に膨大な資料ストックが形成されている。にもかかわらず、動画映像資料は10年前も今も基本的には学術論文の補足資料でしかないし、しかも現物が論文に添付されることはほとんどない。最近、若手の研究者では自作ビデオをCD-ROMにし、論文の添付資料とする試みもなされているが、これを業績としてどう評価するかの一般的ルールは成立していないと思われる。

動画映像資料を学術一次資料として論文に生かすには、なによりもまず引用と参照が常に可能とならねばならない。それには二つの障害がある。第一に、原本すなわち原フィルムの多くはアーカイブとして整理されておらず、原本へのアクセスは必ずしも容易ではない。第二に、映像作品の部分を切り取って資料化することは著作権に抵触する可能性が大きいということである。

ひとつの事例を紹介しよう。トヨタ財団では1982年度から91年度にかけて科学史家の中山茂氏を代表とするチームに「戦後科学技術の社会史に関する総合的研究」というテーマで助成を行い、その成果が95年に『通史日本の科学技術』全5巻（学陽書房）として結実した。この初期の段階で博物館映像学研究所の孝寿聡氏のグループでは産業映画に着目し、これを科学技術史の基礎資料として位置付ける研究を行った。孝寿氏が鋼材倶楽部や都立図書館などから借り出して見た産業映画は100本を超える。筆者もいっしょに見て印象的だったのが、1961年に三菱造船がスポンサーとなり岩波映画製作所が作成した「巨船ネス・サブリン」である。三菱造船が長崎造船所で建造した巨大タンカーの建造過程を記録したもので、芥川也寸志が担当した音楽にも当時の高度経済成長に向う雰囲気がよくあらわれていた。われわれが特に注目したのは、造船ドックの背後の山々がみごとに伐採され山肌の一部がおそらくはドックのために削り取られていたシーンである。当時は環境破壊という概念すらなかった時代で、それゆえ今日的な感覚なら構図として避けるかもしれない背景が無意識に映し撮られていたのである。われわれの当初の思惑は、しかるべきテキスト・クリ

ティークの手順さえ確立できれば、産業映画の中からでも研究上の実証データとして資料化できる部分が切り出せるのではないかということであったが、このワン・シーンはまさにそのような考えに合致するものであった。すなわち、過去に膨大にストックされた動画映像を製作者の意図とは別の文脈で——この例では環境破壊の実例として——再活用することができるはずなのである。

しかし、この研究は結局頓挫する。原フィルムへのアクセスがあまりにも多大な労力を要すること。資料化しようとするのとたんに直面する著作権問題。それに資料評価をしようにもフィルムの製作記録すら残されていないといった状況で、とてもボランティア研究では取り組める課題ではないことが明らかとなったからである。⁽²⁰⁾

だが少なくともわれわれは、フィルムやビデオの成立過程、公表履歴、保存などの問題を扱う「映像書誌学」というべき領域が不可欠であること、それとともに、たとえば、アングルの切り替えと照明の角度等からモンタージュの有無を検証するなど、プロの映像製作者の技法を踏まえた映像資料評価（テキスト・クリティーク）の確立も映像書誌学の一環として急務であることを実感をもって理解した。

⑥……………データベースをめぐって

これまでデジタル資料原本の諸様態ならびにそれを担うメディア、デバイス等の環境要因について見てきた。ここで視点を変えて、デジタル資料が実際に蓄積され継承される「場」としてのデータベースに着目し、そのデータベースの維持管理というきわめて現実的な業務に即してあらためてデジタル化の危うさを検討してみたい。そこから、実作業に伴うヒューマン・ファクターというデジタル書誌学のもうひとつの課題が見えてくるはずである。

ところで最近では、WEB上のファイル・アーカイブのような不定形の蓄積も広義のデータベースと考えられるようになりつつある。なぜなら、データベースのひとつの機能であるデータの検索・参照について見れば、WEB空間の巡回機能をもった強力なサーチエンジンにより、現在すでに、WEB空間自体が巨大なデータベースとなっているとみなして何の不都合もないからである。しかし、データベースのもうひとつの機能である、計数・処理についてはパッケージ・ソフトの中でのみ成り立つものであり、WEB上の不定形のアーカイブには当てはまらない。そこでまずは、ここでいうデータベースを、データベース・アプリケーション・ソフトによって作られたものに限定することにし、事例としてトヨタ財団の業務用データベースを取り上げることにしたい。⁽²¹⁾

(1) トヨタ財団助成金管理データベースの概要

トヨタ財団の助成金管理業務用のデータベースは1990年代はじめから筆者自身が設計・構築し実用に供してきたもので、2002年2月現在、財団設立以来26年分、5650件の助成対象レコードが収録されている。管理画面はカルテ形式で、助成対象者の属性、助成プロジェクトの概要、成果要約などが和英両文で収録され、さらに助成金の送金や書類の提出などが記録されている。また、90年以降の申請レコードについても国内外あわせて1万6000件以上が収録され、毎年度の選考に際してはこのデータベース上で評点、集計などの作業が行われている。入出力フォームなどを除く、

文字データ部分だけで 35 メガバイトあり、これは単行本はほぼ 35 冊に相当する。

財団のオフィスでは 26 台の PC が LAN 上でこのデータベースを参照し、毎日書き込みが行なわれている。そして財団の WEB サイト（www.toyotafound.or.jp）では ASP（Active Server Pages）プログラムによりこのデータベースのコピーが参照され 5650 件の助成対象レコードが検索・表示されるようになっている。

(2) 市場への従属

1990 年に最初の業務処理システムの骨格を作ったのは MS-DOS 上で動く R:Base というリレーショナル・データベース・ソフトによってであった。ほぼ、並行して国産の桐というソフトでも別のパートを作成した。95 年に各パートを Windows 上で動く Access に統合することにし、プログラムはすべて書き直した。この過程で OS の乗り換え、アプリケーション・ソフトの乗り換え、そして各ソフトのバージョンアップと、ありとあらゆる「不条理」を経験してきた。それでも 90 年代前半の乗り換えについては、ある意味で覚悟の上の発展的乗り換えであったと納得している。しかし、マイクロソフトの一人勝ちが決まった 90 年代後半からは、先に述べた市場メカニズムに引きずられた不本意な乗せ換えを強いられている。

直面している問題の一部を紹介しよう。現在、基幹業務は Access97 で動かしている。しかし、パソコン等機器のリース期間は 4 年で、これを過ぎてマシンも幾たびか更新しているうちにプレインストールされた OS はいつの間にか Windows98 以上、Word、Excel といった定番ソフトも Office2000 以上となっている。そこで Access も 2000 にバージョンアップしようと試みたら、なんとこれまでのプログラムにコンパイルエラーが出て自動的に上位変換ができない。察するところ従来のプログラム文法ではルーズに認められていた部分の解釈がより厳密にされたためのエラーと思われる。そこでいったんは上位変換をあきらめ Access2000 のライセンス契約の中でわざわざ Access97 に格下げした状態で動かしてきた。しかし、Access97 などという商品はとくに市場からは消えてなくなっているのである。プログラムの原本としては 1 枚の CD-ROM が手元にあるのみで、なんとも心細い。いまとなっては不本意ながら上位変換に取り組むしかないだろうとあきらめている。

ならば機器のリースをやめて買い取りにしたら？ おそらくはコンデンサーの寿命がつきる 7 年を目途にパタパタと故障がはじまり、ハード、ソフト一式を取り替えるにはバージョンアップの 2, 3 世代分を一挙に飛び越えることとなって、結果はもっと悲惨なものとなるであろう。

データベースは本来、数十年にわたって安定して運用されるべきものであるが、現実にはその基盤となるハード、ソフトのたかだか数年の寿命という大きな制約を抱えている。そして、この制約を乗り越えてデータベースを維持管理するのは単なるオペレーターのレベルの話ではなく、組織そのものの自己同一性にかかわる話なのである。

その点からいえば、実は大学の研究室レベルで運用するデータベースはその多くが研究者一代限りのようなものであり、さらにいえば国立機関においてさえ数十年にわたってこれを維持管理する制度的な基盤が確立しているかというといささか怪しいというのが現状ではないだろうか。

(3)管理運用上の問題

コンピュータのハードは所詮機械であるがゆえに故障の確率は決してゼロにはならない。またソフトも設計者の予想を越える負荷や他のソフトとの競合によりハングすることがままある。それゆえ、データベースの運用にあたっては毎日のデータのバックアップ作業が不可欠である。トヨタ財団のシステムでもデータベースへの依存度が高まるにつれ、このバックアップ作業やあるいはウィルス対策なども含めたシステム監視作業の負荷が確実に増してきた。つまり、デジタル化ということは必然的にこうした裏方の継続的な作業を要求するということなのである。印刷文書は、酸性紙による劣化の問題はあるにせよ、少なくとも書架に収蔵されればそれ以上維持管理に手間がかかることはない。そこが文書資料とデジタル資料の決定的な違いであろう。

次に、日常的なデータ入力、更新について考えてみよう。データベースでは、入力者が一人か複数か、利用者が一人か複数かによって 2×2 のマトリックスが考えられる。現実には、複数が入力し一人が利用する組み合わせはまずあり得ないので、残る3つのタイプに分けられる。研究上の利用という点では、研究者個人が目的をもって一人で入力し一人で活用するタイプのものがもっともデータ精度が高く、情報の質において優れたものになる。次いで、データベースの扱いに優れた研究者が、自分自身の研究に活用したデータを広く複数の人に公開するタイプのものが有用である。逆に、往々にして役に立たないのが、複数の入力者が関与し、利用者も複数であいまいなタイプのものである。なぜ役に立たないかというと、このタイプでは一般に誤字・脱字率が極めて高くなるからである。

トヨタ財団のシステムでは、いったん入力されたデータは、選考委員会用の一覧として出力参照され、さらに理事会用の資料として出力参照され、最後に年次報告書の記述として印字される。したがって実質数回の校正作業を経ていることになり、誤字率は極めて低い。しかし、たとえば某情報研究機関が提供するデータベースでは、データ入力は丸投げの外注で業者にまかされ、せいぜい一校程度を経たところで一括してデータベースに投入されている。このぐらいの校正では数万字に1字程度のミスは避けられないであろう。ちなみにこの程度の誤字率は書籍ならばまず刷り直しである。

実はこの誤字率の問題はデータベースに限らずインターネット上で流通するデジタル文書すべてに通底する大きな問題なのである。多くの人が既に経験として知っているように、印刷刊行を前提としないワープロ文書では誤字・脱字は多くなる。手軽に直せるということもあって緊張感がゆるみ、校正回数が少なくなるためと思われる。ワープロ文書には隣接キーのミスタッチに起因する固有の誤字がある。また、近年OCR（光学文字読取機）が多用されるようになり文字形態の類似に起因する新型の誤字がしばしば見られるようになった。こうした誤字を含むデータが、印刷物として一度も印字されずにデジタル資料として蓄積されるようになったのが今日のWEB上のサイバースペースと呼ばれるものの実態であろう。先に、インターネットの普及は印刷を経由せずに世界にデータを配信する道を開いた、とポジティブな評価を行なった。しかし、その裏には印刷というフィルターを通らずにゴミ情報が垂れ流しとなっているスラムのような光景も見え隠れしているように思う。

ところでトヨタ財団のシステムは複数が入力を行い、複数が利用する。しかし各人が入力した結果が直接自分の業務に影響するため、タイプとしては個人が入力し個人が活用するものにむしろ近く、データの精度は高い。それでも何代かにわたって担当者が代わっていくと送り仮名や文体あるいは外国人名のカナ表記などの表記上の揺れが必然的に拡大していく。これはいわばデータベースにおけるエントロピー増大則のようなものである。こうした表記上の揺れは、統計処理などの作業を通してある程度は補正することができる。いいかえると、統計分析に活用することでデータベースは内部的整合性を維持することができるともいえる。

また、ここでふれた担当者の代替わりという問題は、実はデジタル資料の長期保存という点で最もクリティカルな問題でもある。どんな組織でも一般事務について帳簿や業務マニュアルの引き継ぎはおおよそその手順が決まっていよう。しかし、われわれはデジタル時代に突入してからまだ日も浅く、ハードディスクに格納された目に見えないデータやシステムについて、帳簿や業務マニュアルを引き継ぐような作法を確立しているとは言い難い。その中で、膨大に蓄積されたデジタル資料が、ささいなヒューマン・エラーで瞬時に消滅するという悪夢も、けっして杞憂とかなづけるわけにはいかないのである。

(4) デジタル書誌学からみたデータベース

データベースの内部整合性を維持するためにデータの補正を行なうと述べた。その延長上には過去に入力したデータのミスを訂正することも含まれる。そこでひとつの問題に気づくことになる。

たとえば、レコードを過去に遡って訂正することを想定しよう。過去に間違った記述内容で公開された資料は、書誌学的には、間違っていること自体に意味があることも考えられる。単なる誤字・脱字ではなく、当時には意識されなかった差別的表現が使われていたり、外国人の人名表記法が違っていたりということは、なにがしかの意味を含むものである。そこでもし、データベース上で過去の記述の間違いを訂正するなら、書誌学的にはそれは史料の改竄ということにもなるのである。

これに対処するため、データベースに更新履歴を残すことで訂正の経緯を示すことも考えられるが、レコード単位の更新履歴は可能としてもカラム単位で履歴を残したのではおおよそデータベースとして実用的でなくなるだろう。あるいは日々のバックアップ・ファイルをそのまま凍結保存版として毎日蓄積していくという方策が考えられなくもない。しかしこれも現実的とは思えない。

つまり、書誌学的な観点からデータベースの原本とは何かと問うてみると、データベースの本質である追加更新可能な柔軟性と、書誌学的一義性との間には原理的に相容れないものがあり、データベースには書誌学的な意味での原本は存在しないということが明らかになるのである。

(5) WEBアーカイビング

先に、WEB上のファイル・アーカイブが強力な検索エンジンによってデータベース的な様相を呈するようになってきたと述べた。実は、このような日々生生流転していくWEB上の情報がある日ある時のWEBとして凍結保存するような途方も無い試みがアメリカなどでWEBアーカイビングとして既に行われている。たとえば、アメリカの The Internet Archive (<http://www.archive.org/>) には 2001 年 10 月現在で世界中の約 100 億ページ、100 テラバイト以上の過去のWEBページが保存され

ている。ちなみにアメリカ議会図書館の2000万冊の蔵書の情報量は20テラバイトというからそのWEBアーカイブの巨大さがうかがえる。デジタル化資料の保存という点ではこれこそひとつの極相を示すものであろう。

しかし、こうした凍結保存版もけっしてすべてを網羅するわけではない。そもそもWEB上を巡回して情報を収集するロボットはWEBのバックエンドに隠れたデータベースの中までは侵入できないからである。さらに、100テラバイト以上のデータを溜め込んだハードディスクもいつかは寿命がつきる。それを次の世代に継承していくという本当の困難に直面するのはまさにこれからのことであろう。

おわりに

これまでに述べたことを、デジタル書誌学の課題として少し整理してみよう。第3節では「デジタル資料」の態様を(1)実物のデジタル写像、(2)コンピュータ生成像、(3)文字データ、(4)プログラムそれ自体の4つに大別し、それぞれに、原本の同定やオーセンティシティの保証などをめぐって、文書の書誌学では経験したことのない、新たな課題が生じていることを指摘した。それらの課題に対して、デジタル資料を学問の基礎資料として位置付けていくための長期的にも有効な指針を打ち出していくには、第4節で見たように、コンピュータ技術に関する相当に深い知識と経験が要求されるものと思われる。第5節では映像書誌学を提起したが、既に100年の歴史をもつ映像資料の世界においても、実は、文書の書誌学に相当するような基礎的な取り組みは行われていないといつてよい。肉眼で扱うことができるフィルムにおいてすら資料化の基礎ができていないとすれば、肉眼でとらえることのできないデジタル・データの資料化における基礎づくりの困難さは、生半なものではないと覚悟すべきであろう。第6節では、デジタル資料を実際に蓄積し継承する「場」としてのデータベースについて考察した。そこでの要点は、たんなる技術の問題ではなく、組織、制度のあり方にまでおよぶヒューマン・ファクターの重要性ということであった。

このように見てくると、デジタル書誌学の本当の課題は、いったい誰がそれを担うのかという「人」の問題であるように思われる。かつて、トヨタ財団で、慶応大学斯道文庫の故阿部隆一教授を代表とする「国書漢籍総目録の編纂」という大プロジェクトの一部を助成し、筆者もチームの資料調査に同行させていただいたことがある。そこで知ったのは、書誌学は、山ほどの現物資料と触れることをとおして目利きとなる学問であり、十年かけてようやく一人前となるほどに大変なものであるということだった。デジタル書誌学においても、おそらくこれをこの先100年、1000年にわたる研究資料の継承のための基礎とし取り組んでいくとなれば、とても片手間の仕事ではなくなってくるであろう。デジタル書誌学そのものが課題として広く認識され、それを担う人が出てくること自体がこれからの課題なのかもしれない。

いま、デジタル資料の保存問題を考え始めるきっかけとなったできごとを思い出している。筆者は198X年、助成財団センターで民間助成財団のデータベースを、当時としては最新鋭のIBM5550上でR:Base5000を使って作成していた。財団の助成プログラムは毎年少しずつ改定される。そこで、将来的に改定の軌跡をたどるためには年度ごとのデータベースの凍結保存が必要だろうと思い

立った。当時はまだデータ量もたいしたことがなく、データ分割して5インチ・フロッピーに十分おさまった。それをご丁寧に金庫にしまったものである。そのまま出向期限も終わり忘れて久しかったのだが、199X年、たまたまデータ分析の必要から凍結保存版のことを思い出した。しかし、そのとき既に、IBM5550の5インチのドライブは故障して使われていないという。修理すると数十万で、立派な新品が買えてしまう。もちろん他には5インチのドライブなんぞ無い。かくて外部記憶メディアのはかなさを知ったわけである。

しかし、いまあらためてそれで本当に困ったのかと思い返してみると、実は、それほど困った記憶が無い。つまり、しょせんその程度のデータでしかなかったと、簡単にあきらめがついたのである。

そもそもモノが歴史の時を越えて残るということを考えてみると、案外、なんでもかんでも残そうとすること自体むなしいことのように思えてくる。何かを意図して残そうとする営みは大なり小なり権力への指向と結びついており、歴史家はそのようにしては残らなかったもの、あるいは意図的に消されたものをあえて掘り起こしてくることを仕事にしているようなものである。そういえば、デジタル資料でも例外ではなく、どのみち淘汰されて大半は消えてなくなる運命と割り切ってもいいのではないか。

もしそれでも残したいというのであれば、CD-ROMではなく、ハードディスクである。そしてそのハードディスクを墓守のごとく代々継承していくしか手は無いと思う。そういうと、それは国家事業であるという意見が出てくる。しかし、これまでの歴史を振り返ってみると、実は本当に残ったモノは一子相伝で伝えられたものではなかったのか。つまり、ハードディスクを一子相伝で継承する個人的な営みを通してひっそりと残るべきもののみが残るというのが実相なのではないだろうか。

註

(1)——筆者はトヨタ財団1998年度年次報告の中で「財団のOA化・情報化の歴史」について報告し、その中で「一方、20年の経験からデジタルシステムの脆弱性ということも身に沁みている。デジタルデータは劣化しないといわれるが、機器に依存し、つまり肉眼では見えないため、現実には機器の消滅と運命をともにしているのである。また、複製がいつも簡単のため、どれがオリジナルか分からなくなるという事態は誰もが経験しているはずである。いま時代の趨勢はデジタル化に向かい、大量の古文書や画像資料をデジタル化するプロジェクトが各所ですすめられている。その中で、データの超長期保存や、真贋判定の問題など、いわば『デジタル書誌学』ともいふべき基礎的な領域についてどれほど考慮されているだろうか」と記している。

(2)——「マルチメディア・インターネット事典」Multi-media & Internet Dictionary © Digital Creators Conference

『デジタル書誌学』の項。「……地味に積み重ねられた資料を正確に利用するデジタル書誌学者に受け継がれる……」、「……過去に印刷された眠っている活字をデジタル社会で復活させるという不死鳥のような事業もある」などの記述から著者のいうデジタル書誌学の定義を推測した (<http://www.kaigisho.ne.jp/literacy/midic/data/k19/k19190.htm>)。

(3)——平成13年度文部省科研費特定領域Aとして発足した「我が国の科学技術黎明期資料の体系化に関する調査・研究」(略称「江戸のモノづくり」)では、まさにこのような研究スタイルが目標とされている。

(4)——KODAC社のホームページでは、CDメディアの加齢テストの結果、通常環境で100年、湿度40%、温度25℃の暗所で保管すれば200年もつという結果が出たとある (<http://www.kodak.com/global/en/professional/products/storage/pcd/techInfo/permanence.shtml>)。

(5)——資料の改竄について、「学者の良心」といわれるものが素人の抱く幻想であって、何の担保にもならないということを、最近、実証してくれた考古学者がいた。

(6)——ブリタニカが無料でWEBで引ける。ありがたい時代である (<http://www.britannica.com/>)。

(7)——この記述の前に「図書・記録・文書・地図・図表・書画・写真等を総称するには、これを文献とよべばよろしい」というくだりがある。今日のマルチメディアを示唆するようで興味深い。

(8)——ただし、現在のアカデミズムの中ではマルチメディアにより表現されたプレゼンテーションそれ自体が業績として評価されることはほとんどない。業績の本体はあくまで文字による論文である。筆者はかつて人文系研究者の多くはいわゆるITに乗り遅れているためマルチメディアが業績として評価されないと考えていた。しかし、本稿の趣旨からすれば、紙のメディアにインクで記す「論」の重さは尊重されるべきで、マルチメディアの物量に眩惑されてはいけないという考えも半分はある。

(9)——デジタル化に際しては主に通信回線上の伝送速度の制約から、静止画のJPEG、動画のMPEG、音源のMP3などのデータ圧縮技術が用いられる。その本質は、人間の認識限界をだます技術といえる。

デジタル化手順には、サンプリング周波数×ビット数による標準化、その量子化、さらには符号化という段階がある。量子化誤差、あるいは人間の認識限界外の波長領域の切り捨てなどにより、デジタル化の過程で原本の持つ情報は不可逆的に削減される。

以上、『図解でわかる画像圧縮技術』（1999年、日本実業出版社、越智宏・黒田英夫著）より。

圧縮を伴うデジタル化は今日の技術水準に制約される。それゆえ100年後には、今日の圧縮により捨てられた情報が惜しまれることもないとはいえない。

(10)——コンピュータ上のカラーマネージメントは書誌学的に重要な課題と考えられる。

たとえば産業遺産研究では、レンガの色から焼成温度や成立年代の推定までできるという話がある。その場合、デジタル・データの色の実現性が重要となる。

色の再現をめぐるのはWindowsマシンとMacマシンとで大きな差が出る。また、たとえばデジタルカメラからのプリントの場合、プリンターの違い以前に、マシンの機種差（同じAT互換機でも）により非常に大きな差ができる。

また、ディスプレイ上で見る色（RGB）と、プリントアウトした色（CMYK）とでは差がある。美術品のサイ

トなどでは絵の横にカラーチャートを置いて画像データ化し、補正情報を与えているケースもあり、これは当然のアプローチである。

ただし、「色」は光によって変わると考えれば、レンガの色に基づく推定はあくまで現物を前にしたときの総合的判断の一環として意味をもつものであり、デジタルであろうがアナログであろうが、写真からは導けないものであろう。ここでいいたいのは、少なくとも成立年代などの推定を「表現」するには再現性が問われるということである。

(11)——ここにあげた例は、東京慈恵会医大の高次元医用画像工学研究所（鈴木直樹所長）で「Virtual Anatomia」と名づけた三次元アトラスで実現している (<http://www.jikei.ac.jp/ihdmi/>)。鈴木氏によると、CGがCGのように見えるのはまだコンピュータの力不足でデータ密度が十分ではないだけのことで、早晚CGと現実との見分けはつかなくなっていくだろうという。

(12)——『真贋のはざま デュシェンから遣伝子まで』2001年、西野嘉章編、東京大学総合博物館、p.384。東京大学博物館WEBサイト (<http://www.um.u-tokyo.ac.jp/>)。

(13)——『電子透かし その技術的・経済的・法的・暗号的・倫理的側面』1999年、スコット・モスコウィッツ著、セレンディップ。本書にはGiovanniという名の電子透かし（Digital Watermark）ソフトのデモ版が添付されている。試した限りでは手順が煩雑で大量の画像を扱うには向いていないと思われる。もともとこのソフトは不正コピーを摘発する上で動かね証拠になることが主眼なのであろう。電子透かしを埋め込んでいると表明することでコンテンツの不正コピーを抑止する効果を狙っていると思われる。

(14)——「同版か異版か、初印か次印かは、すべて実物を比較しないと、とんだ誤りをする。川瀬氏の古活字版の研究の図版と京都大学附属図書館の竜龜手鑑の実物を比較したら、巻首の陰刻の二字の間に、実物にはごく細い線が白く入っていた。そこで、少なくともそのページには組み替えがあるものと断じ、京大の本をマイクロに写してもらって、その引き伸ばし焼付をもらったなら、その細い白線は全く消えていた。であるから、マイクロの引き伸ばし焼付は全く信用し難いことを知った。写真を使って比較しようとするなら、原寸大の写真をとらないとだめになることがある」（前掲書p.67）。

(15)——本段落の事実関係の詳細はすべて『図解でわかる文字コードのすべて』清水哲郎著、日本実業出版社、2001年に負っている。

(16)——『コンピュータ帝国の興亡—覇者たちの神話と内幕—』ロバート・X・クリンジリー著、アスキー出版局、1993年、下巻p.237。

(17)——『記憶装置に関する調査報告書—磁気記憶装置(HDD・FDD)・光ディスク関連装置—』2001年3月、社団法人電子情報技術産業協会。

(18)——巨大化したプログラムでは部品化したモジュールを組み合わせて使うようになる。たとえばダイナミック・リンク・ライブラリー(DLL)はこうした部品のようなもので、本体プログラムから参照され部分的な処理を行い、結果を本体に戻す。プログラムのバージョンがあがるとDLLも書き換えられる。しかし、汎用性のあるDLLだと別の古いプログラムからも参照されていたりする。これがある日、別のプログラムの都合で勝手に書き換えられるわけで、古いプログラムはDLLの動作の違いからハングする。この障害を元に戻すには結局、古いプログラムから順番に再インストールしなければならない。こうした状態が俗にDLL地獄といわれる。これはエンドユーザーが垣間見ることのできるコンピュータのブラックボックス的一面である。しかし、Windows NTの開発裏話を記した『闘うプログラマー』(G・パスカル・ザカリー、日経BP出版センター、1994年)によれば、OSの開発者でさえ誰一人としてプログラムの全体像は把握できていなかったという。こうなると、ソフトの中はまさにブラックボックスで、「人知を超える」という表現があながち誇張ではないこともわかっていただろう。

(19)——「ハード、ソフトの急速な発展」とはいわず「変化」といったわけは、たとえばR:Base Proのようにバージョンアップによりかえって機能が低下したソフトも過去にあったからである。

(20)——後に国立科学博物館が産業技術史研究プロジェクトの一環として産業映画を取り上げ、さしあたりの成

果として短編100選が選定されWEB上で紹介されている。ここには彼の「巨船ネス・サブリン」も収録されている。ただし、紹介文だけだが……(<http://sts.kahaku.go.jp/main.htm>)。

ちなみにアメリカでは1000本以上の短編映画を収録しダウンロードできるようにしたWEBサイトが既に存在している(<http://www.archive.org/index.html>)。

(21)——今後、WEB上でXML(eXtensible Markup Language)の規格が一般化していくとWEB上のファイルも構造化が進むと思われる。XMLはいわばパッケージ・ソフトのデータベースの構造定義とデータ内容とをセットでHTML同様のタグ付テキストに吐き出したようなものである。その最大の特徴は、個々のデータベース・アプリケーション・ソフトに依存せずに、ある程度複雑な構造をもったデータを記述することができるという点にある。これによって、学術研究資料をWEB上で共有しデータベース的に累積していくことも技術的にかなり展望が開けてきた。

(22)——外国人名のカナ表記は元来統一の不可能なものでありこれを検索キーにすること自体データベースの設計としてまちがっている。しかし、ならば英語表記はどうかということこれも実は非英語圏の人々の名前を表記するときには著しく揺れが生ずるものなのである。トヨタ財団の助成対象者にはタイ、ラオス、ヴェトナムなどアジア各国の人々がおり、この人たち自身がときによって自分の名前の英語表記を変えてくることなど日常茶飯事である。ならば各国固有文字のUnicodeで入力したら？ 検索利用というデータベース本来の機能が成り立たなくなることは明らかであろう。

(23)——たとえば研究者の所属について統計分析をするには「慶応義塾」「慶応大学」「慶大」などの表記のばらつきが無いほうが好ましいのはいうまでもない。

付録 画像DBツール (Album Tools) について

(1) ツールの基本思想

このツールは画像データベースそれ自体というよりは画像データベースの作成を容易にするための補助的なツール群と位置付けられる。しかし、既存のブラウザーとは異なるこれまでにないカテゴリーのソフトである。ここでの基本的な設計思想を以下に示す。

1 デジタル上の仮原本

デジタル書誌学の考え方に基づき、デジタル化された画像資料を原則的にデジタル上の仮原本とみなす。

この仮原本のファイル形式はネット上での配信を想定してJPEGとする(ただし、高画質とするか標準画質とするかなどは、特に規制しない)。従って、圧縮前のBMP形式画像や、Photo-CD 画像、各種デジカメの特殊規格による画像などは、仮原本以前の資料として別途管理されることが望ましい。ただし、これらの画像には通常、さらに元となる写真や絵画など物理的実態が存在するはずで、これら実物こそが真の一次資料として管理・保存されるべきである。

デジタル化された画像資料には世界中にただひとつというファイルネームをつける。具体的には、作業者名、年月日、シリアルナンバーを組み合わせたロングネームを生成する。ただし、既に名前がつけられており、当面の重複が見込まれない場合はあえてリネームはしない。

仮原本となってから以後はレタッチを行わない。

ファイルコピーではなくショートカット(エイリアス)を活用することで、ひとつの画像ファイルを複数の分類項目に仕訳できるようにする。

2 画像に文字を従属させる

通常、画像データベースと称するものは構造化された文字列(カラム群)に画像ファイルもしくはそのファイル・パスをひとつのカラムデータとして従属させている。

その作成の過程では、文字化された概念枠組み(分類枠)が先にあり、そこにひとつひとつの画像の意味を読み取りながら、画像を配置していく方法がとられる。

しかるに当ツールでは、画像のサムネールをドラッグ・アンド・ドロップで操作することにより、概念枠組みがない状態でも直感的な分類を可能とする。分類した後一括して名前付け(すなわち分類概念の付与、もしくはキーワードの付与)をすることができる。

操作の詳細は3節で述べるが、発想の基本は画像に文字を従属させるところにある。

(2) 実的な利用局面

1 既存データベースと画像データとの関連付け(産業遺産DBの例)

産業遺産データベース(URL <http://www.ied.co.jp/isan/sangyo-isan/JS-TOP.htm>)では7000件の産業遺産レコードがまず整備され、次に別途集積されていた2万点の写真が整理され、このレコードと関連付けられた。

ここでは、物件レコードひとつに対し、多数の写真が存在する場合と、ひとつの写真に多数の物件が写っているため、このひとつの写真を複数の物件レコードが参照する場合とが生じる。

ひとつの写真を複数のレコードに対応させるため、サムネールから、瞬時にショートカットが生成されるようにした。

通常このような作業ではファイルネームの文字列をデータベースのカラムに入力することになるが、これを画像のサムネールをデータベースのカラムにドラッグするだけで入力できるようにし、大量作業を可能とした。

2 画像データへのキーワード付与（産業遺産DBの例）

通常の画像分類では、ひとつひとつの写真に対し思いつく限りのキーワードを与え、これを写真の枚数分だけ繰り返すことになる。その作業過程で、当初の分類概念は徐々にずれていき、さらに大量作業に倦み疲れて、途中で挫折することも少なくない。

ここでは、サムネールをディスプレイ上の仮想空間に自由に配置し、重ね合わせることで、直感的な仕訳を行い、その後にグループ全体に一括して名前付けを行うようにした。これで上記のような作業は大幅に時間が短縮され、しかも概念のずれはなくなる。

この名前付けされたグループは入れ子構造により、上位概念に括り直したり、下位概念に分解したりできる。

こうして、ひとつひとつの写真からみると、そのそれぞれは階層化された複数の分類概念をもつようになり、それは一括してデータベースのテーブルとして吐き出すことができる。

3 大量画像に基づく形態学的分類（考古資料—釣針—分類の例）

国際日本文化研究センターの赤澤威教授の依頼により、数カ所に分散した古代の釣針の写真整理を行った。

ここで、整理された結果から、画像のみの操作による形態学的分類が十分に可能であることがうかがえる。

(3)具体的な操作と特長

1 画像ファイルの変換ツール

このツールの機能は市販の画像ブラウザーなどにもあるもので珍しくはないが、世界でただ一つのファイル名を自動的に生成するなどの特長がある（図1）。

2 画像とレコードとの照合ツール

このツールではディスプレイの半面に画像のサムネールを展開し、もう半面にデータベースのテーブルを表示し、画像サムネールのドラッグによって画像ファイル名がテーブルに入力できる（図2）。

3 直感的画像分類ツール

このツールがいわば Album Tools の中心であり、その特徴は以下に要約される（図3）。

- ①画像のサムネールをディスプレイ上で仮想空間に自由に配置できる
- ②サムネールをグループ化し自由に文字ラベルをつけられる
- ③分類結果をCSV、データテーブルなどの形で一括文字化して吐き出せる



図 1 画像ファイルの変換ツール

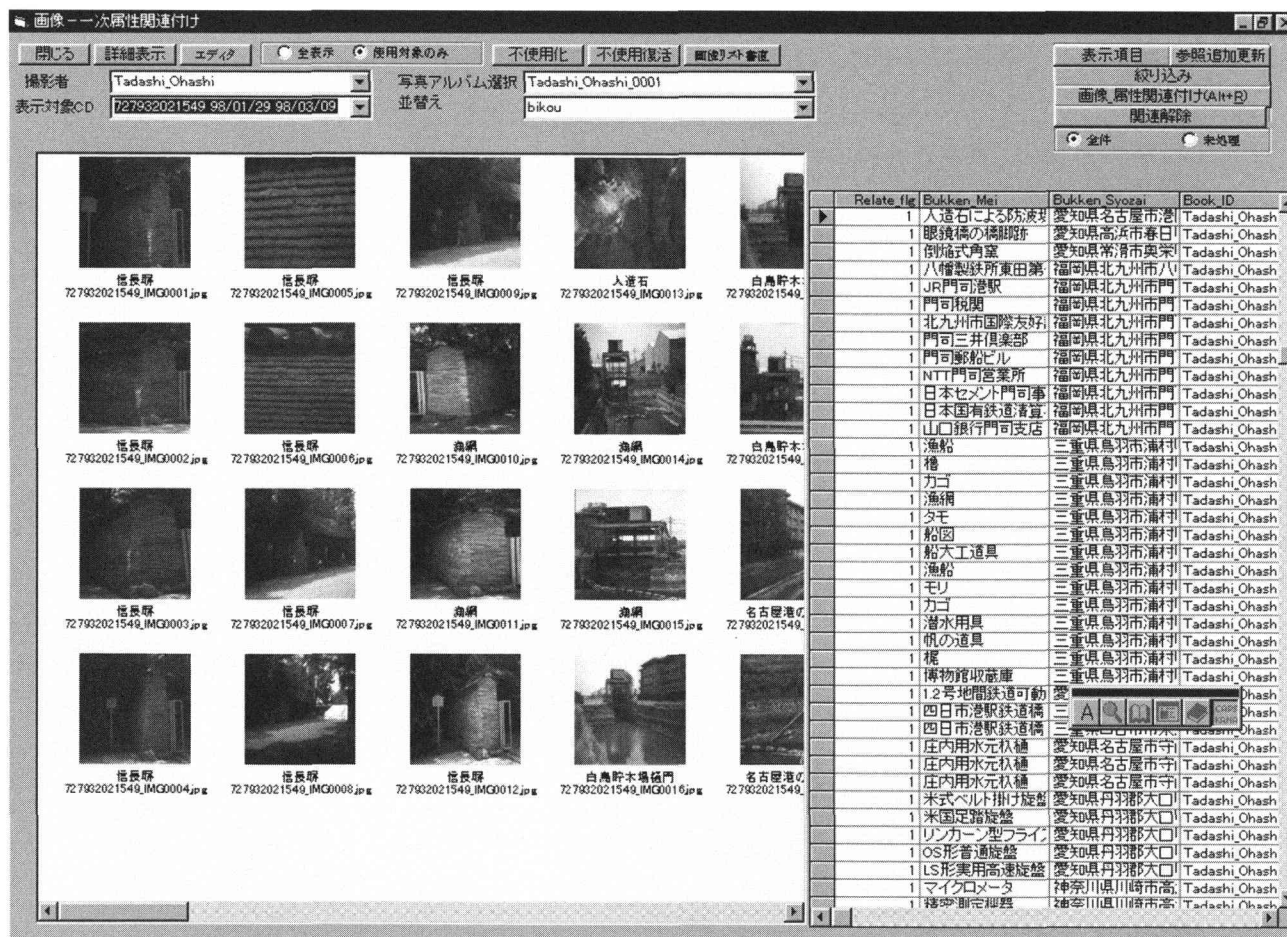


図 2 画像とレコードとの照合ツール

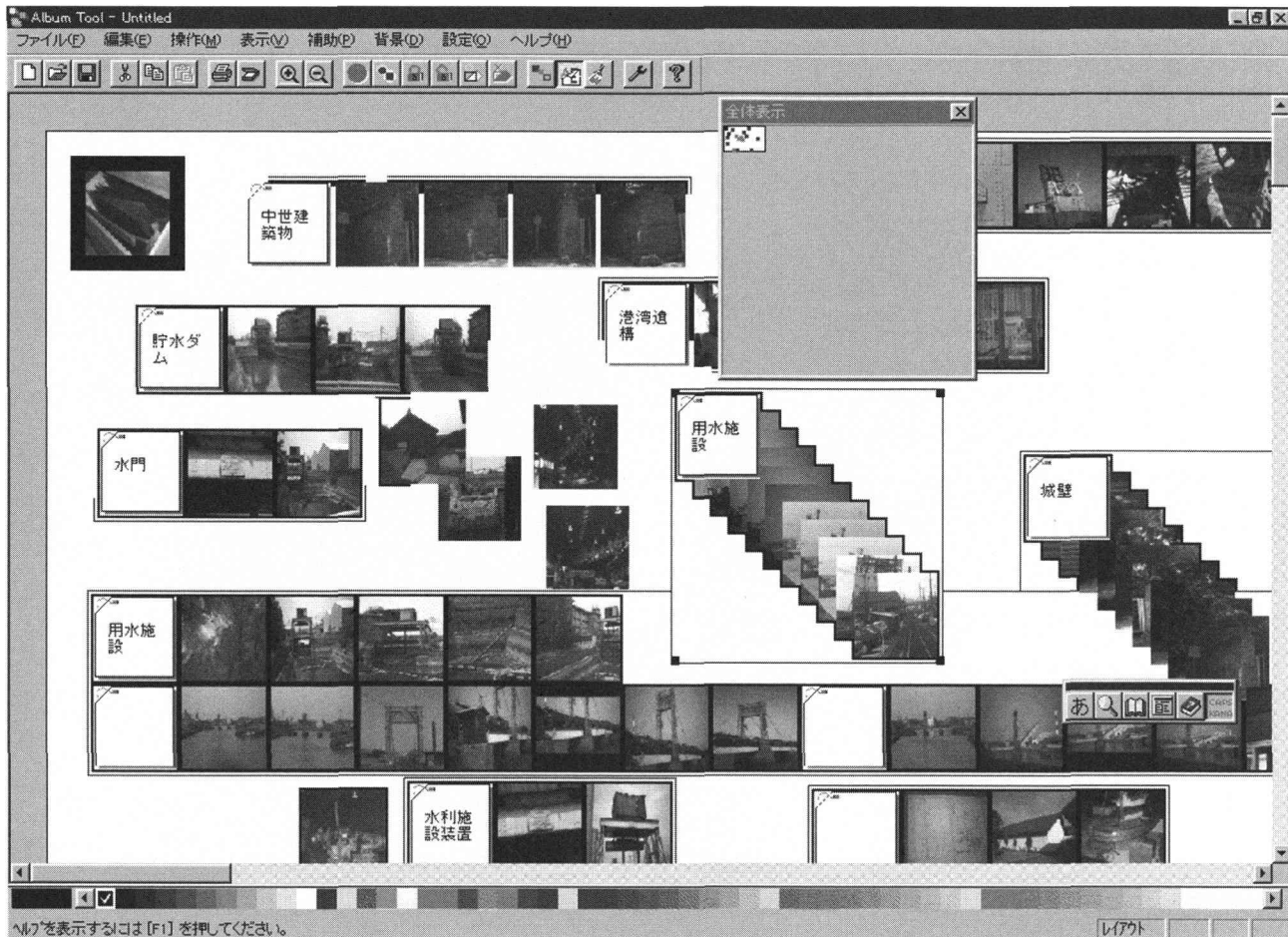


図3 直感的画像分類ツール

(4)欠 点

産業遺産DBの作業にカスタマイズされており、汎用的なユーザーインターフェイスとなっていない。プルダウンメニューで表現できない操作が多数あり、ソフトの習熟に時間がかかる。ユーザーによるβ版テストなどができないため、バグが多く、残念ながらこのままでは一般の利用に供することができない。

(トヨタ財団, 国立歴史民俗博物館共同研究ゲストスピーカー)

(2003年5月16日受理, 2003年7月18日審査終了)

Some Thoughts on “Digital Bibliography”

KUSUMI Masaaki

Digitization of research materials, especially in the field of social science and humanities, promises the development of new kinds of research methods, and new possibilities for collaborative research. But dependency on digitized material may lead to less attention to—and therefore loss of—the original materials that are digitized. And because digital storage of data is dependent on particular hardware and software, and because that hardware and software quickly become obsolete following technological changes and market demands, it is risky to expect digital storage to be long-lasting.

If scholars are to be responsible for introducing new research techniques and methods of access to research data, then they must also accept the responsibility of understanding the current fragility of digital media and attempt to find an proper way of handling digitized materials and data in terms of transmitting it intact to future generations.

To do so, my suggestion is that it will be efficient to apply the analogy of “critical bibliography” to such digital material. I refer to this as a “digital bibliography” in my paper. I try to define digital material from various perspectives and to point out issues of hardware — and software — dependence from a practical point of view.

I also refer to “cinema bibliography,” meaning techniques that should precede the digitalization of moving picture material.

Finally, I describe the various aspects of human error that might occur during the process of digitization by taking an example how the Toyota Foundation currently handles grant database system. I mention, too, a new trend toward Web archiving as an extension of local databases. And, as an appendix, I introduce the “Album Tools” that I created to handle a database of images.