

電子化のススメ

水山昭宏(日本考古学協会)

2000年3月10日に東京都の区市町村埋文担当者研修で「報告書の近未来」と題した発表を行いました。本稿は、その発表要旨をベースに大幅に手直したものです。コンピュータ技術を、単に道具としてではなく、情報(文献)のメディアとして活用しようという提案です。

筆者は、1997年6月の日本語PDF(後述)の登場を踏まえ、1997年度末にPDF主体の報告書の電子化(CD-ROM添付)を実行したのですが、最近ようやく電子メディア化が表立った話題になってきたということでしょうか。2年前の時点でも、技術の仕組みは整っていましたし、既にパソコンの一般的な性能も、メディアとしての使用に耐えるレベルに達していたのですが...

電子化の動機

報告書電子化の動機を一言でいうと、埋文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。従来のように紙と印刷技術に依存したままでは、限界があるということです。

報告書が重い、多い、高い、といった不満(諦観?)は、誰でも抱いているはず。電子化すれば、軽く、安くなり、そしてネットワークの活用で、情報の検索性も高まるはず。研究者は皆、既に膨大な蔵書を抱えて、物理的な困難に陥っているはず。

発行点数は多いにも関わらず、報告書は印刷部数が限られています。研究者でも入手は困難ですが、報告書が市民の目に触れる機会は、なおさら確保し難くなっています。数の問題も、電子化は解決してくれます。CD-ROMなら複製コストが安いということもありますし(20万円で1000枚)、Webを活用すれば部数という概念すら無くなります(というか、無限です)。

ユニバーサルアクセスを分析的に解説するなら、以下のようになります。

- 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
- 2)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の情報デザイン、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

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(提供手段に応じた容量の最適化は必要ですが)。

特にカラー写真の情報は、確かに電子化の威力の一つです。鮮明なカラー写真は、実測図を補完してくれます。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。もし余裕があるなら、3D(スリーディー:立体)表現やGIS(ジーアイエス:地理情報システム、あるいは空間情報システム)もいいでしょう。

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、といった文献サービスの階梯を意味しています(文献本体とは、文献全頁の電子版をさす)。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(オンライン)情報サービスの階梯の一部に位置付けられるべきです。

遺跡地図の電子化も同様で、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、以前にISEKIDAS(Inventory of Sites by Excavation and Kinds of Implements ; Documents for Archaeology Service)として発表したことがあります(水山 1997)。これは3階層の構成を考えています。

第1層:遺跡インデックス...いわゆる遺跡台帳・地図。GISも指向する

第2層:遺跡の詳細データベース...遺跡・遺構・遺物の项目的な詳細情報

第3層:文献サービス...報告書、及びその他のワイド&リッチな情報

ISEKIDASは文化のアーカイブを目指したもので、研究者や文化財関係者等のプロフェッショナルをメインターゲットとします。Webなら、市民向けの楽しいコンテンツも必要と思われるでしょうが、それはISEKIDASのしっかりした幹から咲く花と考えてみて下さい。

報告書電子化の基本

報告書を電子化していく上での基本要件を4点挙げます。

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(テキスト主体でよい)
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+拡張版)

報告書電子化に関わるフォーマットは、以上に加えて、CADデータの提供もありえます(標準フォーマットが確定できないのが実情ですが)。ポストスクリプト系の画像は、PDFから抽出

できます。また色々な意味で、将来はXML(あるいはXHTML)の活用が有望です。

この基本要件の特徴は、PDFをメインとし、HTMLやその他の標準的なフォーマットを補助的に組み合わせることです。機能的なデータベースの組み込みは、とりあえず見送っています。そこまで余裕がないということもありますが、現状でのデータベースは、ランタイムをインストールすることになり、プラットフォームも選ぶことになります。報告書電子化にあたっては、なるべく業界標準フォーマットのみ採用し、クロスプラットフォームを目指すべきです。

PDF の特徴

PDFとはPortable Document Format(直訳すれば、可搬文書形式)、つまり別のコンピュータで読み込んで、元のレイアウトを崩さずに表示できる電子文書形式のことです。PDF閲覧のためには、無償配付・再配付自由のAcrobat Readerを使うわけですが、製品版のAcrobat(約25000円)でもOKです。製品版は、PDF変換・編集ができるようになっています。

電子文書とは、ワープロやDTPの画面での表示状態(印刷プレビュー)と同じことですが、PDFの場合はスタンダードを目指したものですから、AcrobatないしAcrobat Readerが作動する限り、元の作成アプリケーションが無くても、文書をオリジナル通り表示できます。Acrobatは、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。テキストや2値画像のスムージング(アンチエイリアス表示)も快適です。

PDF自体の制作(変換)は自動的に行なえるので、殆どコストはかかりません。この点こそPDFの最大の特徴かもしれません。PDFはコンピュータのプリントを電子的に横取りして定着したようなフォーマットなので、印刷用のデータ編集が、そのまま電子文書の制作過程になってしまうわけです。PDFによって初めて、書物が、紙から離脱する可能性が開けたといえます。

PDFは、印刷されるはずの文書と、内容的にもレイアウト的にも、データの的にも同一といってよいフォーマットです。画像の解像度を印刷用データと同じにしておけば、そのまま商用印刷も(理論的には)可能ですから、印刷データのバックアップにもなります。通常の閲覧用PDFの解像度でも、レーザープリンタでの出力には適しています。それでいて、ディスプレイ上で終始する電子文書としても最適化されています。

PDF の発注

PDFの制作は、印刷発注の時に同時に頼んでしまいます(仕様書に含めておきます)。自前でDTP(DeskTop Publishing)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は印刷直前のデータから作る必要があります。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、但し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(とりあえず、目次を「しおり」に反映させる程度で充分です)。

以下にPDFの仕様例を示します。解像度別(容量別)に3種の制作を薦めます。これは文献サービスの階梯性を考慮するため、オンライン版は現状の通信基盤で公開するためのものです。

	高解像度版	閲覧版	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。どちらも、無くてもとりあえず困らない機能です(本来は重要な機能ですが)。現状での基本的な閲覧版は、3.0互換にしておくべきでしょう。

閲覧環境は、やはり4.0の方が快適です(2値画像の表示が滑らかです)。なお、4.0のインストールにあたっては、3.0をアンインストールしてからにしてください(市販のアプリケーションやドライバには、何時のまにか3.0をインストールしてしまう場合もあるので注意)。

HTML の役割

電子文書なら、HTMLはどうでしょうか。図版や写真を含めて、印刷のために作成したレイアウトを再現するためには、HTMLは向かないようです。(さらに問題なのは)レイアウト再現性を度外視しても構わないのですが、HTML版に画像を含めようとする場合、HTMLの作動環境に最適化した軽い容量に手直しする必要があります。素材を用意したり加工する二度手間を考えると、HTML版への画像収録は、躊躇してしまいます。DTPソフトの書き出し機能を利用するとしても、何しろ遺跡の報告書の図版は巨大で膨大です。余程単純なレイアウトでないと、うまくいかないようです。

一方、これからの電子時代には、紙上のデザインから離れた、電子時代に相応しいデザインも求められているはずで、そもそもWeb(インターネット)の隆盛なくして、今日のコンピュータ/デジタル氾濫の状況もなかったわけで、Webとの連続性を考えても、HTMLは重要です。HTMLは、電子文書規格の古典たるSGMLのアプリケーションですから、素性は確かです。SGMLはXMLに移行しつつありますが、この種のML系フォーマットは、PDF以上に研究に値する存在です。

結局、現段階での報告書電子化に於ては、テキスト主体のHTML版を作成することにしまし

た。図版を含めた完全版は、PDF でよとしました。

デジタルアーカイブ

アーカイブとは、文書や資料を、文化遺産として永年保存していくことです。報告書の電子化は、当然デジタルアーカイブへの収録を期しています。またWebを活用することで、全国(全世界)から、24時間、制約のない文献閲覧を可能にすべきです。無論、きちんと書誌情報や要旨等を用意し、利用しやすいものにすべきです。一言で言えば、電子図書館です。

実は保存ということを考えると、デジタル化は必然です。過去の映像作品、音楽作品等、続々とデジタル媒体に移されつつあります。緻密性も品位も保存性も、デジタルが有利だからです。無論、絵画や遺物のような現物の保存は、何時までも重要に決まっていますが、文章や図版・写真のような情報は、映画や音楽と同様に、人類の文化遺産としてデジタル化されるべきです。

写真のデジタル化

例えば、写真のデジタル化は、情報量の点で、アナログ写真とどのように比較できるでしょうか。35mmフィルムの品位は、印刷すると、概ね葉書大が限度であることが分ります。それだと、印刷に必要な画素数は、横 1500 ~ 2000 程度です。余裕をみても 1920 × 1280 あれば充分です。これだと、ベタのRGB 画像で7.2MB ですから、最も控えめにJPEG 圧縮すれば2 ~ 3MB になります。1/3 程度のJPEG 圧縮なら、画質の劣化は全く見分けられない程、高品位なものです。つまり、約2.4MB あれば、35mm 相当の写真はデジタルアーカイブできるということです。これなら CD に 250 コマくらいは収録できます。

ブローニーや4 × 5 ではどうでしょうか。これらの解像度は、35mm フィルムの縦横2 ~ 3 倍程度ですから、フォトCD でいうと16BASE と64BASE の中間くらい、画素数でいうと4096 × 3072 あれば充分ではないでしょうか。これだと、ベタで36MB ですから、圧縮して12MB 程度になります。これでも、CD に50 コマくらいは収録できます。

デジタル情報の保存性をカバーするのは、現在ではCD-R です。CD-R とは、記録対応のCD の規格のことです。容量や形状はCD と同じです。追記はできますが、書換えは出来ないので、公文書の保存にも向いています。メディアはブランド品で1枚160円程度です。一応、寿命は100年とされています。100年以内には、次の媒体に移した方がよさそうですが、数十年後には現在では考えつかないような記録手段が開発されているでしょう。

写真のデジタル化には、一般的にはフォトCD が有効です。CD-R 1枚に100コマの収録ですが、コストは1コマあたり150 ~ 170円くらいです。遺跡調査に関わる写真の大半は、現在はフィルム上にあるはずですから、一刻も早くフォトCD化を進めるべきです。フォトCDにしておけば、退色の心配もなくなります。フィルムの退色を防ぐには低温が決め手なのですが、大掛かり

なことになりますし、頑張っても、結局退色を遅らせることしかできません。

デジタルカメラも、300万画素に達していますから、上述の35mmフィルムの品位に迫っています。画素数だけが品位を決めるわけではありませんが、民生用でも上位機種なら、35mm一眼レフの代替になりえます(ここぞという時の中・大型カメラの出番は残ります)。デジタルカメラなら、コスト的な撮影数の制約もなくなります。建設省では、既に工事写真のデジタル化を推進しています。埋文でも写真のデジタル化を本格的に推進すべきです。

Webの推進

行政電子化の流れも重要です。国は、はっきりと各種の電子化推進を打ち出しています。Webの活用も当然要求されています。国の文書を見ると、電子政府(電子統治: Electronic Governance)だの、バーチャル・エージェンシーだの、ものすごいキーワードが並んでいます。埋文も行政の一部なら、そうした行政電子化の流れと無縁ではいられないはずで

す。埋文関連の機関、報告書の発行主体は、すべからくWebを立ち上げるべきです。つまり公式サイトです。その動機は、本稿で述べたように、埋文自身の中から発するべきです。

ちなみに学協会・研究会・同人会も電子化を推進すべきです。会報や論文集は、電子ジャーナル化すべきです。メディアはWeb中心で、場合によってはCD-ROMも組み合わせることになるでしょう。やはり研究者にとって、発表やコミュニケーションの場が電子化していかないと、考古学全体の情報化は完結しないと思われます。こうした考えから、筆者は「研究会」公式サイト構築も推進しています。その一例として、江戸遺跡研究会公式サイトを紹介しておきます。徐々にですが、会報の電子化と公開も進めています。

江戸遺跡研究会公式サイト [Web] <http://www.ao.jpn.org/edo/>

PDF主体のCD-ROM 報告書例

都立学校遺跡調査会 1998 『岡本前耕地遺跡』

西国分寺地区遺跡調査会 1999 『日影山遺跡・東山道武蔵路』

都内第二遺跡調査会 1999 『西台後藤田遺跡』

加東郡教育委員会(兵庫県) 1999 『埋蔵文化財調査年報1998年度』

加東郡教育委員会(兵庫県) 2000 『東古瀬・坊ノ下遺跡/東古瀬・坊ノ上遺跡』

2000年3月発行予定は、東京都管轄分だけで5例あります。

関連文献・URL

水山昭宏 1997 「報告書の電子化 - 考古学及び埋文関連文書の電子化と公開について - 『考古学ジャーナル』

418, 1997年6月号, pp.37-39. <http://www.amy.hi-ho.ne.jp/mizuy/9706.htm>

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