

古文書データベースの試み

嶋本 隆一¹⁾、干場 悟²⁾

はじめに

「安政の飛越地震」は、当館の総合研究のテーマである。歴史分野として、この地震災害にかかわる資料（古文書等）を調査・収集・整理・解説・分析し、これらを企画展や研究紀要などに発表している。この活動の中で特に資料収集と収集された資料の閲覧と展示に関し、新たに獲得した古文書のデジタル化の手法と今後の課題について述べていきたい。

1. 過去を振り返って

当館として膨大な近世資料を収集対象としているため、年間数千単位のコマ数のフィルム（ポジ・ネガフィルム）が必要であることは開館以前から予想された。しかし、限られた調査費で、効率の向上を図るためには、フィルム不用のデジタルカメラ（以下カメラ）の活用は、不可欠であると考えた。

デジタルカメラを使用した第1回目の調査は、1998年金沢市立玉川図書館の近世資料室で行われた。当初のカメラ（1996製）の性能（35万画素）は低く、普及も限られていた。光線の具合で出来ばえが左右され、一定の明度と鮮明さを持った画像にするには自然光が最適であることはいうまでもなかったが、長時間安定した場所を探すことはほぼ不可能であった。最近2000年の自動調整機能がつき画素数も上がった（200万画素数以上）デジタルカメラでは、少々暗い場所（図書室内など）でも、鮮明で一定の画質を確保することが可能となっている。

撮影された画像は、コンピュータ（以下パソコン）に取り込み、サムネイル表示で一覧性が高まり、自由自在に目的の画像に移動できた。ネガやポジフィルムでは、プリントにしなければならず、数が多いと一覧することが難しく、目的の写真に行きつくまでに時間がかかってしまう。最初のデジタルカメラは35万画素であるから、最大画像サイズは（640×480ドット）と

いう規格だった。これを拡大すると画像がかえって読みづらくなってしまった。しかし、デジタル化への一歩となったという手応えを得て、この作業は継続することになった。

翌1998年、画素数130万というカメラを確保し、画質は飛躍的に向上した。拡大画像も鮮明となり、閲覧可能となった。着脱可能なメディア（この時、スマートメディアを使用）を介してカメラからパソコンへの取り込みは、一回のメディア交換で撮影枚数が50枚（後には200枚以上）程度となり、効率のよい作業となった。書状などの横長の文書は分割して撮影し、ファイリングソフトを利用し、画像を重ね合わせて1枚に編集することもできるようになった。しかし、高解像度で撮影すればするほど当然ファイルのサイズが大きくなり、パソコンの上での動作が不安定になった。解像度を上げて撮影すると鮮明な画像になるが、最新の性能を持ったハードウェアが必要になってくる。カメラの性能は向上したものの、その画像を調整し閲覧するパソコンの性能を高める必要を強く感じるようになった。文書の画像をどのように利用するかを年頭に、カメラの解像度を調整しながらの撮影となった。当館では、徐々にパソコンの性能があがり、データの読み込み速度が高まり、また同時に200万画素数のカメラの利用がはじまった。そのため、高画質と自負していた一昨年昨年の画像は、急激に色あせてしまい撮影のやり直しとなった。撮影時には画像の規格を最大限にすれば、パソコンの性能が不十分でもいずれ利用度が増すことは容易に考えられる。そのため、原画像を他の場所に保管しバックアップを取った上で、画質やサイズを落としたものを閲覧用としてパソコン上で展開するという方法を採用した。

最近のインターネットはADSL方式などの登場で各段に速度が速くなった。以前9600bpsという速度で電子メールの送受信をしていたことと比較すると、

¹⁾立山カルデラ砂防博物館

ADSL方式では8Mbps、1,000倍程度向上したことになる。パソコン本体のCPUもペンティアム100Mヘルツから1,000M（1ギガ）ヘルツという数字に大きく変化し、外部記憶媒体としてのフロッピーディスク（1.44Mバイト）はCD-R（700Mバイト）に置き換わり、ハードディスクの容量も1Gバイトから40Gバイト以上というように、各分野で飛躍的に性能が向上している。また、データを加工するソフトの種類が増加し、年々製品がバージョンアップしている。このような状況をふまえて、現在利用できなくても、将来確実に利用可能という判断もあり、現場での撮影は大きめの解像度で撮影した。

調査活動（画像収集）を重ねるごとに、さまざまな工夫を考えた。手ぶれ防止対策としてリモートコントロールで撮影し、実寸を明示するため定規を接写台に配置した。事前に準備した付箋を画像に取り込むことで、撮影後の整理を円滑に行うなど様々なノウハウが身についた。文書の保存状態を含めた全体像を把握するため、調査館保存用の大封筒そのもの、そして封を開け内部資料を並べた画像などもデータベースとして貴重なものとなった。従来より利用されていた「やさしくファイリング」というソフトは一般的ではなく、他のパソコン上では該当するソフトがインストールされていないと作動しない。どのパソコンでも自由に閲覧できる方法がないか、次の課題となった。汎用性のあるフォーマットとしてアドビー社の「アクロバット」を採用したのは平成12年であった。撮影した画像をアドビー社の「フォトショップ」で色調補正し、文書ごとにファイル名を付しPDFファイルとして保存することで汎用性の問題は解決した。大きな画像を取り込んだ結果、PDFファイルのサイズは大きくなったが、前述のパソコンの進化によって文書の閲覧には大きな支障はなく、原本に近いもの、それ以上の鮮明さで展開することができるようになった。

画像数、古文書点数が増加するにともなって原画像の点数が増加し、サイズも膨大なものになり、パソコン本体のHD内で収容しきれなくなった。したがって、これらのデータをどのようにして保存していくか検討しなければならなくなった。作成済みのPDFファイルの総量は12GBある。現在これらを60GBの外付けHDに保存し、バックアップとしてCD-Rに保存している。文書の枚数によって一枚のCD（700Mバイト）では収納しきれず、1つのPDFファイルを分割して保存して

いる。今後はDVD-Rなどの採用で分割バックアップを回避したいと考えている。

2. デジタル化の効果

デジタル化をすることにより、データが劣化することがなく、記憶媒体に永久保存が可能となる。また、ネガ・ポジフィルムや紙焼きに比べると温度・湿度に対して神経質になる必要がなくなる。ネガ・ポジフィルムをスキャナーでパソコンに取り込み、デジタル化することで、保存することも可能となる。また、拡大・縮小そして文字の濃淡調節が可能のため、解読しやすくなる場合が多い。さらに、画像データとして様々な加工がしやすくなり、展示としても有効に利用が可能となる。たとえば、国絵図など非常に大きいものは難しいが、一畳程度の絵図や一紙文書などは、分割撮影により合成が可能となった。

3. デジタル化の手法

3-2 撮影機器の準備

撮影はライティングに時間がかかり、長時間となると文書の水分を蒸発させて変形する場合があるため、ハロゲンライトなどは使用しない。内蔵ストロボも、文書中央が明るく周辺部は暗くなってむらとなるため使用しない。室内の窓側でできるだけ光線のむらのない場所を選択する。接写台にはスケール（紙製）を配置し、必要に応じて付箋を準備する。同時に大量の画像を撮影する場合、パソコンへの取り込み作業を迅速に行うためデータを保存するフォルダーを作成する。リモコンを利用することにより、手ぶれを防ぐことができる。接写台に文書を設置する係と撮影する係、最低二人必要となる。

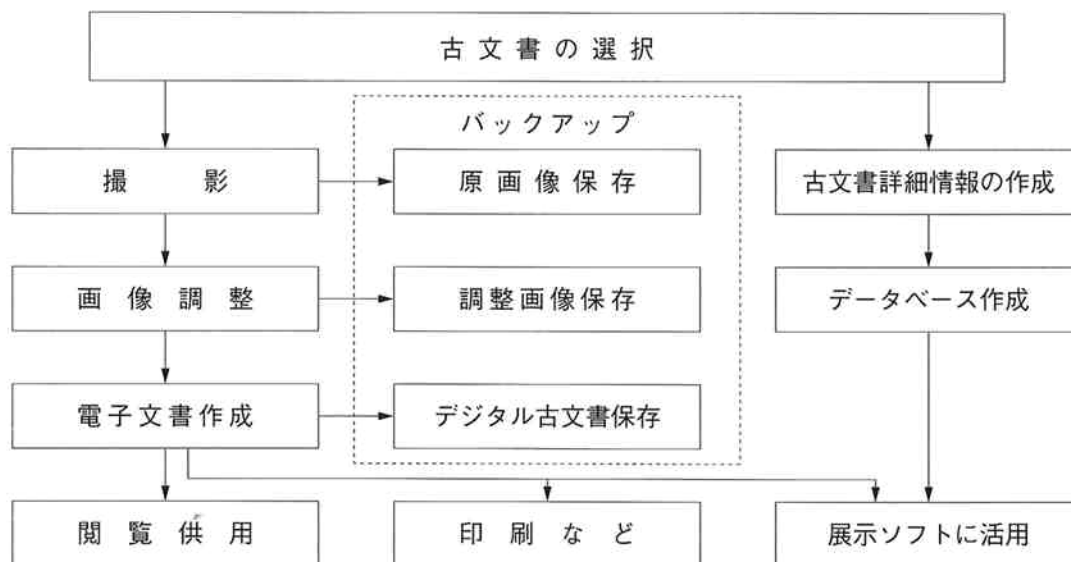
文書周囲に余白を残して撮影すれば、レンズの収束によるゆがみは少なくなる。特に接写を全画面で行うと四辺が湾曲し、実際と異なるようになる。約20%は余白を残し、画像処理の際に切り取る。

撮影前にカメラの時刻あわせとファイル名の設定であり、オリンパス社の場合はファイル名をオートすれば、年月を最初の2桁に埋め込んでしまえば連番になり、重複することはない。

3-3 画像サイズの決定

撮影する場合、カメラの解像度を高くすると鮮明な画像となる反面、ファイルサイズが大きくなりデータ処理に時間がかかるようになる。B5サイズ程度の冊

3-1 基本的な流れ



子を見開きで撮影する場合、通称HQ（ハイクォリティ）モードでの撮影で十分であり、この場合圧縮されたJPG画像ファイル500キロバイト程度になる。不要な部分を切り抜きすることで、実際には200キロバイトから300キロバイトのサイズになり処理しやすくなる。質の良い画像を小さく変更しても鮮明さは失われないが、反対に小さなサイズの画像を大きくすることはできない。画質の良い画像を撮影し、必要に応じてサイズを縮小して使用することが必要である。

3-4 画像の調整

Ulead社のPhotoexplorを利用すればExif情報を保持したままで、ファイル名の変更・画像切り取り・調整・変換が可能である。Photoshopで画像調整をすると、カメラ情報は失われるので注意が必要。Photoshop利用の場合、切り取りサイズを固定すれば、マウスのクリックで一定のサイズの画像を確保できる。最近のカメラの性能は飛躍的に向上し、色調補正、明度補正などはほとんど不要になった。不要な部分を切り取って画像サイズを小さくし、ファイルを小さくする必要がある。

3-5 バックアップ

通常デジタルカメラの撮影はJPG（圧縮ファイル）形式で保存される。最近のカメラでは、Exif情報が記載され、写真情報の閲覧が容易になった。200万画素のカメラをHGモードで撮影した場合、JPG画像ファイルのサイズは500キロバイト程度になるから、CD一枚に1400ページが保存できる。調整後の画像もJPG形式で保存しておくことが便利である。ファイル名を同じにし

てCDに保存し、VCDケースに読み込ませると、ファイルの格納位置やサイズなどがメディア交換なしで一覧できる。ファイルが破損した場合、どのCDに記録したのか保存場所を瞬時に検索できる。さらに、このファイルをCSV形式で入力し、エクセルで加工すればファイルの合計数などを一挙に計算できる。

最近のCD-Rドライブは、バーンプルーフなど書き込みエラー防止対策がついているが、異なった書き込みソフトを同一の機種にインストールすると、正常に作動しなくなるので、避けるべきである。

完成したアクロバット形式のファイルは、通常利用しているHDのクラッシュに備えてバックアップが必要となる。全文書の確実なバックアップは外部記憶媒体（CD-ROM）が便利である。基本的にバックアップは、それぞれの過程で外部記憶媒体に保管しました。撮影直後のカメラから取り込んだ画像、それを切り取って画像補正を行った調整済みの画像、それを完成したアクロバット文書の三種類のバックアップを行った。

3-6 デジタル文書化

アクロバットへの取り込みは、アクロバットリーダーでは作動しない。本体のアクロバットそのものが必要となる。操作はマニュアルどおりで簡単な操作となっている。完成した文書にサムネイルを作成し、必要に応じてパスワード設定も可能である。PDF化をする場合、一度に取り込みできる画像数は50枚以上となっているから、ページ数の多い場合は何度かに分けて取り込みが必要になる。完成したアクロバットファイルは別

途CDに入力してバックアップが必要となる。

ワードやエクセルのファイルをアクロバット形式に変換した場合、そのサイズは軽くなるが、画像ファイルをそのまま取り込んだ場合は、あくまでもイメージファイルとして読み込むのでサイズは軽くない。JPG圧縮で50キロバイトの画像をPDFに変換すると10倍の500キロバイト程度になる。100枚の画像をPDF化すると50メガバイトになる。CDに保存すれば、CDそのものが数冊の古文書という感覚で利用することもできる。

3-7 機材

- ・デジタルカメラ
(オリンパスC-2020Z・リコーDC-4・カシオQV-100)
- ・モバイル用パーソナルコンピュータ
(シャープ製メビウスPC-PJ2)
- ・接写台

3-8 小道具類

- ・定規 (透明プラスチック製をコピーしたもの。)
- ・付箋紙
- ・筆記用具
- ・手袋 (薄いもの。)
- ・メンディングテープ

3-9 関連ソフト

3-10 絵図の撮影

冊子文書や一紙文書のみならず古絵図のデジタル化は技術的に難しい作業であるが、分割撮影し、つなぎ

合わせることで一つの大きなファイルにまとめて保存することができる。画像をつなぎ合わせるソフトも多くあり、境界線も目立たなくなる。画像のサイズが大きくなりますが、拡大したものを鮮明にディスプレイ上で表示することができる。マウス操作で拡大、移動が可能である。画像サイズ、明るさなどを同一条件にし、撮影は重複部分を15%程度にすることで成功する。

4. 今後の課題

頁数の多い文書は中綴箇所が盛りあがる。横長の一紙文書は、折って保存されているため、山折り谷折りの部分がカメラと文書との距離を複雑にするため、鮮明でないところがある。文書に物理的な力を加えずに、平坦にして撮影する方法について課題が残る。

また、文書の整理番号をどのように分類するかが課題である。データベース化とデジタル化は平行して進めなければならない。単に文書をデジタル化したのでは、閲覧が容易であるというメリットでしかなく、膨大な数の資料に統一された整理番号を付し、資料の情報を与えることによって汎用性が高まる。資料の情報としては、文書作成の年代・作成者(著者)・頁数・法量(形態)・様式・内容の概要・所蔵場所・保管状態などがある。また、ファイル情報は、ファイル数・ファイルサイズ・作成月日(撮影月日)・作成元などがある。これらの詳細情報が、他の機関と統合された一定の書式を持つならばさらに利用しやすくなる。

(2001年3月1日脱稿)

5. 参考資料

デジタル化資料一覧 (一部抜粋)

番号	名称	サイズ メガバイト	作成年月日	備考
PDF01	魚津御用言上留第2巻.	417	平成12年10月6日	
PDF02	魚津御用言上留第1巻.	450	平成12年10月8日	
PDF03	寄合御用留(安政五年).	564	平成12年10月8日	
PDF04	諸事御用留.	162	平成12年10月8日	
PDF05	大地震山抜御達書写.	76	平成12年10月8日	
PDF06	カルデラ博.	73	平成12年10月14日	
PDF07	菊地文書(御用留).	21	平成12年10月8日	
PDF08	名称未設定絵図.	53	平成12年10月8日	
PDF09	郷土博物館.	47	平成12年10月14日	
PDF10	県立図書館.	173	平成12年10月14日	

古文書データベースの試み

番 号	名 称	サ イ ズ メガバイト	作 成 年 月 日	備 考
PDF11	立山博物館.	88	平成12年10月14日	
PDF12	□×ヲ.	101	平成12年10月13日	杉本文書
PDF13	□□キ.	93	平成12年10月13日	杉本文書
PDF14	□□ケ.	160	平成12年10月13日	杉本文書
PDF15	□□コ.	32	平成12年10月13日	杉本文書
PDF16	□□ミ.	185	平成12年10月13日	杉本文書
PDF17	赤□へ.	143	平成12年10月13日	杉本文書
PDF18	赤△ノ.	169	平成12年10月13日	杉本文書
PDF19	赤▲へ.	204	平成12年10月13日	杉本文書
PDF20	赤○○ニ.	102	平成12年10月13日	杉本文書
PDF21	赤丸○○イ.	108	平成12年10月14日	杉本文書
PDF22	地水見分録.	209	平成12年10月13日	杉本文書
PDF23	木村立嶽.	70	平成12年10月13日	杉本文書
PDF24	赤○+ユ.	160	平成12年10月13日	杉本文書
PDF25	赤○+口.	305	平成12年10月13日	杉本文書
PDF26	赤○○イ.	108	平成12年10月13日	杉本文書
PDF27	赤○+ハ1.	329	平成12年10月14日	杉本文書
PDF28	赤○+ハ2.	329	平成12年10月13日	杉本文書
PDF29	赤□ヒ.	526	平成12年10月13日	杉本文書
PDF30	赤□セ.	451	平成12年10月13日	杉本文書
PDF31	赤□ソ.	53	平成12年10月14日	杉本文書
PDF32	赤○+オ.	528	平成12年10月14日	杉本文書
PDF33	赤△井.	23	平成12年10月13日	杉本文書
PDF34	赤○+サ.	52	平成12年10月13日	杉本文書
PDF35	赤△レ.	606	平成12年10月13日	杉本文書
PDF36	赤▲ホ.	355	平成12年10月13日	杉本文書
PDF37	赤△ユ.	36	平成12年10月13日	杉本文書
PDF38	赤▲リ.	54	平成12年10月13日	杉本文書
PDF39	赤△ワ.	729	平成12年10月13日	杉本文書
PDF40	杉本文書□□のル_CD4.	458	平成12年10月11日	杉本文書
PDF41	□ケ.	776	平成12年10月13日	杉本文書
PDF42	杉本文書□□のル_CD3.	566	平成12年10月11日	杉本文書
PDF43	杉本文書□□のル_CD2.	555	平成12年10月11日	杉本文書
PDF44	杉本文書□□のル_CD1.	559	平成12年10月11日	杉本文書
PDF45	□□リ.	510	平成12年10月13日	杉本文書
PDF46	○+オ.	126	平成12年10月13日	杉本文書
PDF47	△△ホ5.	296	平成12年10月13日	杉本文書
PDF48	○○セ.	153	平成12年10月13日	杉本文書
PDF49	△△ホ2.	229	平成12年10月13日	杉本文書
PDF50	△△ホ1.	100	平成12年10月13日	杉本文書
PDF51	△△ホ3.	116	平成12年10月13日	杉本文書
PDF52	△△ホ4.	198	平成12年10月13日	杉本文書
	合 計	12,985		

