

資料デジタル化事業の方法と成果

1. はじめに

画像資料の「活用」「資料化」を表題に掲げる本プロジェクトの中核となるのが資料デジタル化事業である。これは、資料の持つ情報をデジタルデータという形に置き換えることで、保管・加工・活用を容易にすることを目的とするものである。

以下、5年間の作業を通じて蓄積したノウハウの概要と、新たに認識された課題について述べたい。

2. 対象資料

本事業で対象としたのは、大場磐雄、柴田常恵、折口信夫、櫻井満、杉山林継、宮地直一の各氏が収集した資料のうちの画像資料である。この画像資料には写真(原板・プリント)、拓本、絵画、スケッチ・実測図等が含まれる。

各コレクションの内容については次節以降で詳しく紹介されるが、資料の多くは各氏の研究者・行政官としての職務に伴って集められた記録資料であること、各コレクションとも千点単位の量を持つものであり、また、多様な学問領域に渡っていることなど、共通の特徴が認められる。こうした特徴は、従来のデジタルアーカイブの多くが国宝・重要文化財級の資料あるいは特定少数の一括資料を対象としたものであったのは大きく異なる。また、図書のように目録規則が定められているものや、土器石器や民具、文書などのように博物館で一次資料とされるものとも異なる。デジタル化による資料化にあたっては、これらの点に留意した方法を確立する必要がある。

3. デジタル化の方法

画像資料は画像そのものの部分と、撮影対象・撮影者・撮影年代・サイズ・材質・付されたメモ類などの部分とに分けることができる。デジタル情報としては、前者は画像データ、後者はテキストデータとして管理される。プロジェクト開始当初は、この両者を一括して入力・検索するシステムを構築したが⁽¹⁾、その後の実際に作業を進める過程で、作業効率やデータ管理の容易さの点から入力作業段階では両者は分離して行う方法に改めた。

作業はコレクション毎に行なったため、その作業内容の詳細についても次節以降コレクション毎に述べられるが、ここでは一般化して作業工程の概要を示したい(図1)。

画像については、クリーニングその他必要な処置を行なった後、整理番号をふる。これは各コレクション、あるいはその中での一括資料毎に与えられた略号(ob、os、sj等)と、その中での通し番号を組み合わせたものである。次に、スキャナーもしくはデジタルカメラを用いた入力作業を行なう⁽²⁾。ここで得た高画質のデータは外部ハードディスク、DVDなどに保存する。さらに、サイズを落としたデータを公開用に作成する。なお、カラーリバーサルフィルムについては外部委託し、複写とフォトCD化を行なった。

テキストデータについては、資料に残されたメモ類の読解や、サイズ計測・材質確認などを行なう。撮影年代や撮影場所などの項目に関しては、資料そのものから得られる情報が少ない場合、他の資料を参照して項目を埋める必要があるが、資料性の観点からこの両者は区別すべきであろう⁽³⁾。

実際の作業にあたっては、どの部分を採用するか、どのように表現するかなど判断に迷う場合も少なくないため、作業を複数の人数で行う場合は資料数による分担よりも項目毎に分担することが、作業場の一貫性を保つ点で効果的である。また、旧字や記号・色などデジタルデータとして不向きな情

報も記録する必要があるため、紙媒体の台帳を作成した上で、デジタル入力を行うことが望ましい。入力はコピー・ペースト、並び替え、抽出、自動入力などが可能で、しかも一般に普及していることから市販の表計算ソフトを用いた。

なお、この2つの工程は画像データの方を先に行う方が効果的である。資料に残されたメモ類が少ない場合、テキストデータに関する作業は画像データを手かがりにして行わなければならないが、その際、既に画像のデジタル化が完了していればオリジナル資料が受ける負担は少なくなる。

また、作業は基本的に資料デジタル化研究室で行なったが、この他にも折口資料については文学部付置研究所である折口博士記念古代研究所で、また、Web運営を含む事務作業は日本文化研究所第5研究室でも行なった。これらは、学内基幹LANで結ばれ、データのやり取りが可能である(図2)。

4. 資料の公表

このようにしてデジタル化した画像データおよびテキストデータを統合し、Webサイト、写真資料目録、データベースの3つの形で資料の公表を行なった。

Webサイトでは、大場磐雄博士写真資料、折口信夫博士写真資料、櫻井満博士写真資料、杉山林継博士写真資料の一部を公開した。いずれも全体の公開ではなく、コレクション内でも比較的まとまった量の資料群について、試験公開として公表したものである。現状は、サムネイル画像による一覧から、解像度150dpi・最大長700ピクセルの画像データへのリンクを張る形式であり、収録データ量・システムとも低いレベルであるが、いち早く掲載できる点、修正が容易である点、カラー写真の掲載に制限がない点、そして何よりもアクセスのしやすさの点から重要な意味を持っている。

紙媒体の写真資料目録としては『柴田常恵写真資料目録』を刊行した。文字の検索や並び替えなどはデジタルデータの方が勝れているが、紙媒体の利点としては一覧性の高さや書き込みのしやすさなどが挙げられる。また、長期保存を考慮し、中性紙ベースのコート紙を使用している。デジタル媒体と併用し、それとは異なった形で保存・活用する方法として位置づけることができる。

データベースについては、上記2種類の媒体の利点を活かす形で設計を行なった。本書に添付した「大場磐雄博士写真資料データベース」がそれである。市販のソフトを元に設計しているため汎用性も高いものになっているが、キーワード検索とともに、ページをめくるように順番に閲覧することも可能な仕様としているほか、気づいた点などを書き込む場として利用者用のメモ欄も設けている。また、パスワード入力による管理者モードでは、テキストデータの書き換えやその一括エクスポート、公開・非公開の設定などが可能となっている。なお、画像は解像度1000dpi・最大長1000ピクセルとし、これらを別フォルダに格納し、データベースから参照する構成となっている。

5. 成果と課題

以上のように、5年間の資料デジタル化事業によって、大量の画像資料の整理・公開システムをある程度確立することができた。

整理作業のうち、画像データについては、デジタル化の機器や保存形式・保存媒体など理工系の技術的な点で課題を残すものの⁽⁴⁾、概ね当初の予定通りの成果をあげることができた。

問題は、テキストデータの整理である。人文科学において資料として用いる場合、画像データとともに、いつ・どこで・何を記録したのか、というテキストデータは不可欠である。しかし今回扱った画像資料にはしばしばこの記述が欠けていた。また、今回のような膨大な量を抱える資料群の場合、適切な検索機能を備えることは本来必須であるが、画像内容をテキストに変換するにあたっては視点によって様々な描写が可能であるが、何を「撮影対象」として抽出するかという点については今後の

課題である。また、他の人文系資料と同様、「分類」の問題も課題として残されている。

成果公表については、一度のデジタル化で複数の媒体での公表が可能になるというデジタルデータの特性を活かし、複数の媒体での公表を行い、それぞれの媒体の性質にあった用いられ方がなされている。技術的な方法もある程度ノウハウを蓄積し、デジタル媒体については、今後の Web を含めたデータベースシステム構築への手がかりが得られている。

一方、何をいかに公表するか、という点については課題が残されている。特に神社・寺院所蔵資料や民俗資料などについては宗教上・倫理上の問題を検討しなければならない。

このように、整理段階・公表段階の両段階において、技術的には一定の成果が得られたものの、新たに資料論的な、あるいは運用面での課題が浮き彫りになってきた。これらはまさに人文科学的な研究分野であり、既存の学問分野にとらわれず、本プロジェクトの他の部門と連携しながら、学際的・産学協同的な視野から解決すべき課題といえよう。

(中村耕作)

註

- 1) 平成11年度事業報告で報告したもの。
- 2) デジタルカメラよりもスキャナーを用いることの教育的効果については山内利秋「文化財系画像資料の保存から活用へ - 國學院大學での取り組みを例として - 」(日本写真学会主催 平成15年度 画像保存セミナー講演原稿 <http://www2.kokugakuin.ac.jp/frontier/research/yamauchi03.html> 参照。
- 3) この点については、プロジェクトが後半に差し掛かった段階で認識されたため、必ずしも完全に実施できた訳ではない。
- 4) 保存の問題では、ファイルの破損という自体が発生し、DVD と外付けハードディスクで複数のバックアップデータを作成した。

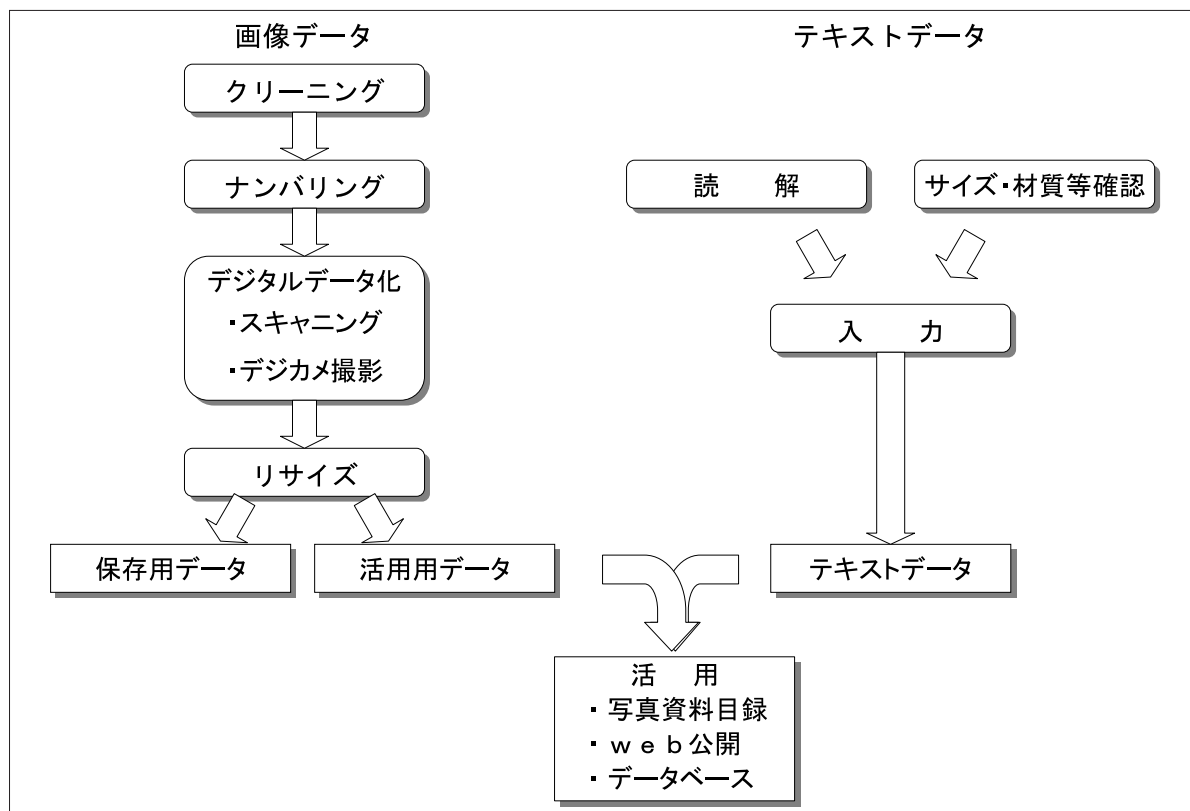


図1 写真資料のデジタル化の流れ

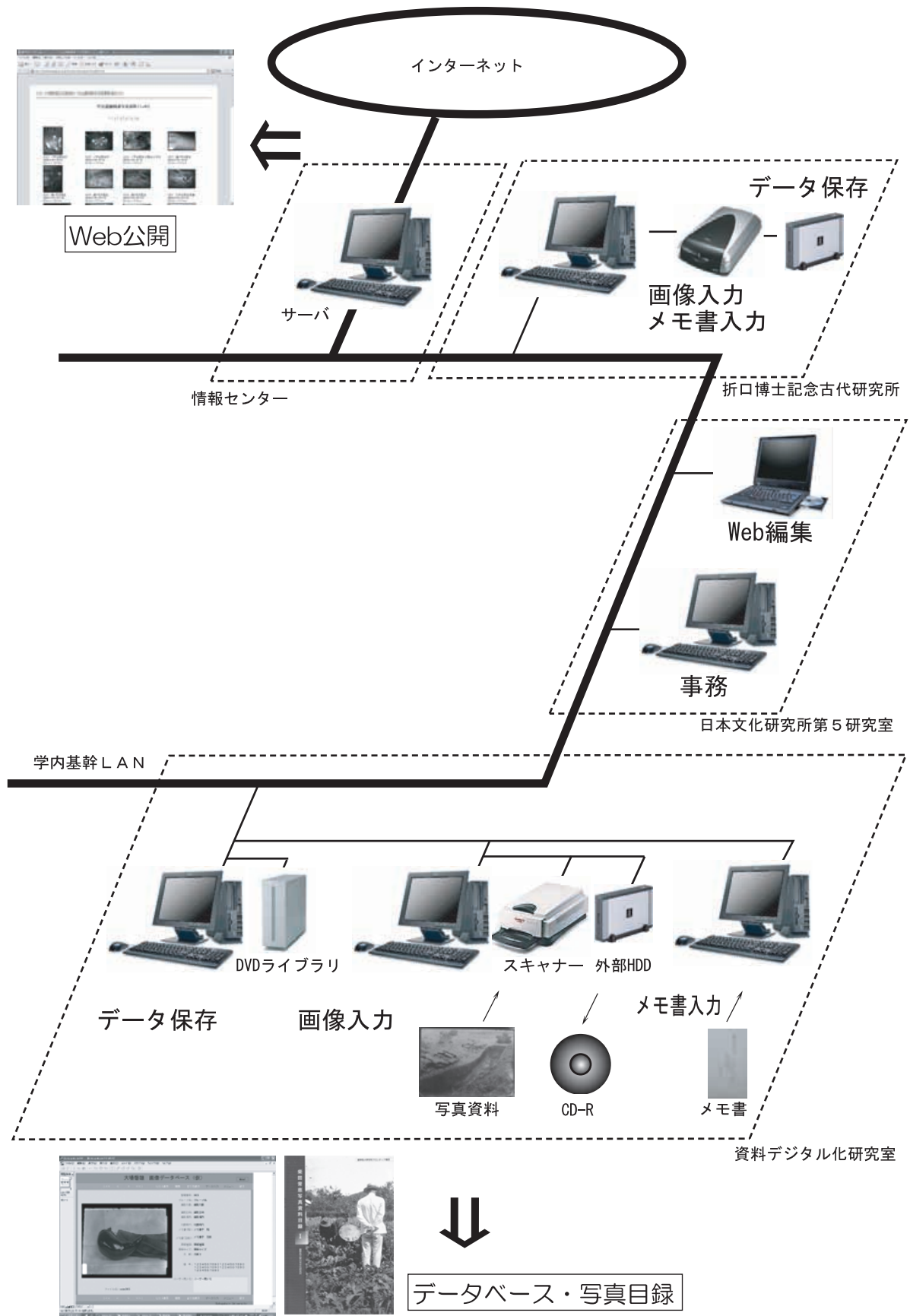


図2 作業環境概念図