

「マンロー関係資料デジタル化プロジェクト」におけるデジタル化作業の過程

The Process of Digitization Work in
“The Project for Digitizing Materials Related to Neil Gordon Munro”
SHIROISHI Rina

城石梨奈

1 はじめに

マンロー関係資料デジタル化プロジェクト（以下、「本プロジェクト」とする）は、日本、イギリスおよびスコットランドに存在している N. G. マンローに関連する資料をデジタルの形態に変換することによってデータとして一所に集積し、それを個々の目的に応じた利用のために提供することを一つの目的としている。本稿では、本プロジェクトのなかで筆者が関わった資料についてのデジタル化の過程をまとめたうえで、デジタル化作業を通して再認識された文化資源のデジタル化の利点、さらに問題点や課題などについて指摘する。

2 デジタル化対象資料の特徴

本プロジェクトに参画したマンロー関係資料所蔵機関は、国立歴史民俗博物館（以下、歴博とする）、北海道開拓記念館（以下、開拓記念館とする）、Royal Anthropological Institute（イギリス。以下、RAI とする）、National Museums of Scotland（スコットランド。以下、NMS とする）の4 機関である。表1 は各機関に所蔵されている資料点数を示しているが、点数だけでなく、種別、大きさ、劣化の程度、既に行われている調査の内容などにおいて多岐にわたっている。デジタル化作業においては、対象資料の特性に応じた機材選定が必要となることから、本節ではデジタル化において注意すべきことという観点から、対象資料の特徴について述べる。なお文書資料については、複数枚に渡る論文や手紙資料は、1 枚ずつをそれぞれ1 点として数えている。

歴博所蔵の資料は、ガラス乾板写真とニトロセルロースフィルム写真という、デリケートな資料であり、割れや乳剤の剝離などといった劣化が進行している状況であった。しかし、後述するように既にデジタル化されたデータが存在したため、本プロジェクト内ではデジタル画像の取得は行わず、既存のデジタルデータをデータベースや他機関資料との比較などに活用した。

RAI 所蔵の資料は、原稿や手紙やメモといった文書類、手紙が入っていた封筒、論文などに添えられている厚紙のカード、新聞の切り抜き、紙焼きプリントの写真、写真付きポストカード、ネガフィルム、ガラス製ランタンスライドとそれが入っている箱、大型の地図が含まれる。原稿や手

紙は薄い紙にタイプライターで打たれたものが多く、そこに手書きで書き込みがある場合もあった。うら面にも記入がある場合や、メモ書きのある紙の小片が付箋として文書にのり付けされているものもあった。資料のサイズ、紙の色や字の濃淡などは一様ではなかった。メモやカードには手書きで文字が書かれており、写真資料の多くにもうら面に手書きのキャプションが書かれていた。資料にはすべて RAI によって資料番号が付されており、資料自体にも鉛筆書きで書き込まれていた。それぞれの資料についての資料情報も既にリスト化して整理されている状態であった。

北海道開拓記念館所蔵の資料は、ガラス乾板製のランタンスライドが 100 点である。資料の大きさはいずれも同一で、おもて面には同じ内容のキャプションが貼付けられていた。資料番号や資料情報は整理が済んでいた。

NMS 所蔵の資料は、原稿や手紙などの文書資料、封筒と新聞切り抜きが 1 点ずつ、および紙焼きプリントの写真である。文書資料は RAI と同じく、タイプライター打ちのもので、事前打ち合わせ段階では、紐で複数枚が綴じられている状態であった。写真資料には何点かうら面に手書きキャプションのあるものがあった。プリントされた紙が反ってしまっており、強い圧力がかかると資料が傷むことが懸念された。なお、資料には NMS によって正式な資料番号が付されておらず、資料情報の整理も済んでいない状態であった。

3 各機関所蔵資料のデジタル化作業

3.1 作業の概要

基本的な作業の流れとその手段、それぞれの作業を行った理由を表 2 に示す。準備プロセスでは、対象資料の特徴を考慮して撮影機材や PC 及びソフトウェアを選定する。この際、必要に応じて対象資料と画質的に等価なサンプルを用いてデジタル化のテストを行う。デジタル画像の取得プロセスでは、デジタルカメラあるいはスキャナを用いて資料をデジタル画像化する。画像処理プロセスでは、撮影画像の回転、ホワイトバランス調整や露出補正を行っている。デジタル化された画像のフォーマットが RAW 画像の場合には、RAW 画像を汎用画像フォーマットに変換する、いわゆる現像ソフトを用いて TIFF 画像と JPEG 画像に変換している。RAW 画像とは、デジタルカメラの製造会社が個別に定めている専用画像フォーマットであり、詳細な撮影情報が画像データに含まれているため、RAW 画像を現像する際にパラメータを変更することで、後処理として画像補正を行うことが可能となる。しかし、将来にわたって現像ソフトを運用可能な状態に維持しなければならないという問題がある。TIFF 画像と JPEG 画像は汎用画像フォーマットとして広く普及しており、現状のパソコン環境では追加のソフトウェアを必要とすることなく閲覧可能であるが、画像処理を繰り返し実行すると画質劣化の要因となる。TIFF 画像には非圧縮形式と可逆圧縮形式の両方があるが、どちらにおいても画像に含まれているエッジ情報に損失はないため、文字や線画の記録に適している。一方、JPEG 画像では非可逆圧縮が行われており、データ容量が TIFF よりも少なくても、その代償として画像に本来含まれているはずのエッジ情報が損失している。画像ファイルのリネームプロセスでは、スキャナあるいはデジタルカメラが生成する画像ファイル名を意図的に変更している。この変更する際のルールをネーミングルールと呼ぶが、基本的には所蔵機関の資料番号を踏襲することとしている。データのバックアッププロセスでは、取得した各種デジタルデー

表1 各所蔵機関とデジタル化資料

所蔵機関	デジタル化対象資料総数	資料内訳
歴博	390点	ガラス乾板写真 323点 ニトロセルロースフィルム写真 67点
RAI	2103点	文書 1350点 封筒 46点 写真 670点 ポストカード 1点 ネガ 11点 カード 8点 新聞の切り抜き 9点 箱 1点 ランタンスライド 6点 地図 1点
開拓記念館	100点	ガラス乾板製ランタンスライド 100点
NMS	1626点	文書 1453点 封筒 1点 新聞記事 1点 写真 171点

表2 デジタル化プロセスの概要

プロセス	作業の内容	手段	作業の目的
①準備	機材の選定と用意、テストスキニング／撮影、リハーサル、データ容量や作業効率の見積もり、役割分担の割り振り	資料の類似物と実際の機材を用いての一連の作業のシミュレーションを行った。	作業現場に赴いてから、作業が完了しなかったり、機材とデータのバックアップの容量が足りないという事態にならないようにするためには、十分な計算とリハーサルが必要になる。
②デジタル画像の取得	対象資料のスキニングおよびデジタルカメラ撮影	卓上型カラーイメージスキャナおよび3種類のデジタルカメラを用いた。	画像取得の方法として対象資料の性質や状態に適していることと、作業の効率性、データ容量、デジタル化後の画像の活用のし易さを考慮し、選択して行った。
③画像処理	画像の回転処理、ホワイトバランスの調整や露出補正など、およびRAW画像からTIFF画像、JPEG画像というフォーマットへの現像	使用したデジタルカメラ専用の現像ソフトウェアやその他の画像処理ソフトウェアを用いた。	画像処理は、画像から資料の情報を得るうえで最適な状態にするために、各資料に適したものを適用した。ファイルフォーマットの変換は、将来にわたって使用可能なフォーマットである可能性の担保のため。
④画像ファイルのリネーム	定義したネーミングルールに従って画像ファイル名を変更	資料名と画像とを対応させ、PCのコマンドの一括処理によってリネームした。	画像取得段階においては、取得装置によって自動的に付される画像名がつくが、これでは同じ取得装置を用いた場合に同じファイル名がついてしまう可能性がある。また、どの資料のデジタル画像であるのかを、ファイル名によって対応づけられるようにするため。
⑤データのバックアップ	作業中のデータ、および作業後のデータをそれぞれの方針に従って、PC本体と外付けのハードディスクに複数保存	手動で、またはコンピュータのコマンドを用いてデータをコピーした。	作業中は万一のデータの損失に備え、基本的にそれぞれの段階のすべての画像を保存するようにした。作業後は、複数のフォーマットで、物理的に分散させてバックアップを保持することで、安全性と再生可能性の担保を図った。

タのバックアップを行っている。

それぞれの所蔵機関によって資料の性質や状態、作業の環境等が異なったために、その都度最適な手段を選択しつつデジタル化を行った。したがって、デジタル化に用いる機材や方法、データに加えた処置なども機関ごとに異なっている。本節の以下では、表2の各プロセスについて、実際の作業に即してより詳細に述べることとする。また、各所蔵機関に特殊な点についても触れる。

3.2 デジタル画像の取得

デジタル画像の取得は、歴博所蔵資料以外は、スキャナまたはデジタルカメラを用いて行った。歴博資料のデジタル化については、他機関資料とは状況が大きく異なるため、3.2.3 節において個別に述べる。画像取得の作業は、表3の日程で、対象資料の所蔵機関にプロジェクトの歴博メンバーが赴いて実施した。使用する装置や解像度などの設定は、歴博と資料所蔵機関とが相談のうえ、決定した。

表3 画像取得の年月日

所蔵機関	年月日	作業人数
歴博	1992年 プリント版の作成 2003年 デジタル化	外部委託
RAI	2006年12月6日～16日(実質作業日：8日間)	常時稼働スタッフ3名 一時的なスキャニングスタッフ2名
開拓記念館	2007年6月13日～14日(2日間)	常時稼働スタッフ3名、所蔵機関の資料担当1名
NMS	2008年1月6日～18日(実質作業日：9日間)	常時稼働スタッフ4名、所蔵機関の資料担当1名

3.2.1 スキャニング

スキャニングによって画像を取得したのは、RAI 所蔵の写真資料および文書資料のみである。写真資料にはプリント写真、写真付きポストカード、ネガおよびランタンスライドを含み、文書資料は、文書（手紙および論文）、封筒、カード、地図、新聞の切り抜きといったものである。写真資料と文書資料のスキャニングには、仕様の異なる2機種3台のスキャナを用いて行った。機種の違いによってスキャナ1、スキャナ2と区別することとする。異なる機種を用いた理由は、高解像度が求められる写真資料と、可読であることが求められる文書資料とでスキャナを使い分けることで、各資料に適した画質のデジタル画像を限られた作業期間で効率的に取得するためである。使用したスキャナの仕様は表4の通りである。

表4 スキャナの仕様

仕様項目	写真資料用 スキャナ1	文書資料用 スキャナ2
形式	卓上型カラーイメージスキャナ	卓上型カラーイメージスキャナ
原稿サイズ	A4/US レターサイズ	A4/US レターサイズ
最大有効領域	216×297 mm (8.5×11.7 インチ)	216×297 mm (8.5×11.7 インチ)
解像度	主走査：6400 dpi・4800 dpi	主走査：3200 dpi
	副走査：9600 dpi	副走査：9600 dpi
読み取り階調	RGB 各色 16 bit 入出力	RGB 各色 16 bit 入出力

適用するスキャン解像度は、所蔵機関である RAI が要求する 300 dpi を文書資料のスキャニングに適用したが、写真資料はより細部まで確認できることが理想的であるということで歴博と所蔵機関とが合意し、800 dpi に設定した。

資料のうら面については、写真資料の場合は、何も書かれていない場合でもスキャニングを行った。これは、RAI の写真資料には写真に関する情報がうら面に記入されているものが多くあったため、記入がない資料についても、記入がないということが確認できるようにするためである。文

書資料については、うら面の印刷や書き込みなどがある場合にのみ行った。また、資料とデジタル画像を対応づけるために資料番号を画像中に含ませる必要があるが、RAI 所蔵資料では図 1 の右上にあるように、既に資料番号が資料自体に書き込まれていたため、改めて資料番号札を付すことはしなかった。

以上のように、RAI 所蔵資料はまずスキャニングによってデジタル画像を取得した。しかしながら、作業の過程でスキャン可能サイズを上回る文書資料が多く存在することが明らかとなり、一つの資料を縦や横に分割してスキャニングを行う必要が生じた。このように分割された状態では、継続的な画像データの保存と活用の面で不都合があると判断し、文書資料については全てのスキャニングが終了した後に、記録解像度が低下することは承知の上で、資料全体が一画面に収まるようにデジタルカメラをセットして撮影を行った。

図1 RAI のスキャニング画像の例(文書資料)

スキャナによる資料のデジタル化は、単純なスキャナ操作によって行うことができること、スキャニング場所を暗所にする必要がないこと、資料の平面性を確保しやすいことなどのメリットがあるが、デジタルカメラによる撮影に比して時間を要すること、資料を平面に保つために僅かながらでも圧力をかける必要があること、スキャン可能なサイズに物理的な限界があること等のデメリットもある。各所蔵機関のデジタル化対象資料数と、作業可能な時間、そして活用に当たってのデジタル画像の画質とを考え合わせた結果、他の所蔵機関の資料はすべてデジタルカメラでの撮影という方法を選択した。

3.2.2 デジタルカメラによる撮影

歴博資料と RAI の写真資料以外は、デジタルカメラによって資料を撮影した。資料の種類や状況に応じて異なるデジタルカメラを使用しており、その仕様と撮影対象資料は表 5 に示す通りである。デジタルカメラ 1～3 はすべて異なるメーカーの製品であり、その使い分けはそれぞれの資料の性質にとって最適と思われるものとした。

表5 デジタルカメラの仕様および対象資料

		デジタルカメラ1	デジタルカメラ2	デジタルカメラ3
対象資料		文書資料(RAI, NMS)	RAI の地図, 箱, ランタンスライド	開拓記念館の全資料(ランタンスライド), NMS の写真資料
仕様	撮像素子(CCD)	23.6×15.8 mm	36.86×36.86mm	49.1×36.8 mm
	有効画素数	1,020 万画素	1,600 万画素	3,900 万画素
	ISO 感度	100-1600	100-400	50-800

RAI の文書資料のうら面の撮影については、スキャニングの場合と同様、記入が有る場合のみ行った。RAI の場合は、資料番号が資料自体に書き込まれていたため、写っている資料を特定するた

めに資料番号札等を付す必要はなかったが、うら面の場合は、撮影時に「うら」と書いた紙片を資料の下辺に挿入することで、うら面の画像であることを判別できるように配慮した。

開拓記念館所蔵のランタンスライド資料は、スライドマウント部分に記載されている情報も取得するため、各資料に対して、フロントライティングでおもて面とうら面を撮影し、さらに資料背後からのバックライトによる透過光撮影を行った。撮影された資料の例を図2に示す。透過光撮影では、スライドの内容が見えやすいように、バックライトの余白部分には紙製マスクを用いている。また、いずれの撮影においても、資料の下部にトレーシングペーパーに印刷した資料番号札を付している。また、記録解像度は300 dpiに統一している。

NMSの資料については、文書資料と写真資料のいずれにおいても、うら面は記入がある場合のみ撮影した。資料番号を書いた札および、うら面の場合には赤い印を資料下部に挿入した。図3にその例を示す。また、テスト撮影とこれまでのデジタル化の経験をふまえ、文書資料はデジタルカメラ1、写真資料はデジタルカメラ3によって300 dpiの解像度を目安に撮影した。

スキャニングとデジタルカメラ撮影によって取得した各機関の画像データの画素数の目安は表6に示す通りである。

図2 開拓記念館の撮影画像の例(左から透過光, おもて面, うら面の撮影画像)

表6 各機関画像データの画素数

所蔵機関	画素数
歴博	3072 pixel×2048 pixel
RAI	スキャンデータ(文書)：3395 pixel×2543 pixel
	スキャンデータ(写真)：3234 pixel×3421 pixel
	撮影データ(文書)：3872 pixel×2592 pixel
	撮影データ(スライドなど)：4072 pixel×4072 pixel
開拓記念館	5428 pixel×7230 pixel
NMS	写真資料：5428 pixel×7230 pixel
	文書資料：3872 pixel×2592 pixel

図3 NMSの撮影画像の例(左がおもて面, 右がうら面)

3.2.3 歴博資料の画像取得について

歴博所蔵資料は、本プロジェクトの一環としてデジタル化されたものではなく、本プロジェクトが開始される前に既にデジタル化されていたため、そのデジタル化作業は他の資料とは大きく異なっている。本プロジェクト内で再度デジタル化を行わなかったのは、複数回の作業による資料の劣化を懸念したこと、既存デジタルデータが本プロジェクトの使用目的に十分耐えうると判断したことが理由である。歴博所蔵資料のデジタル化は、外部業者への委託によって行われたものであり、以下にデジタル化の経過を簡潔に述べる。

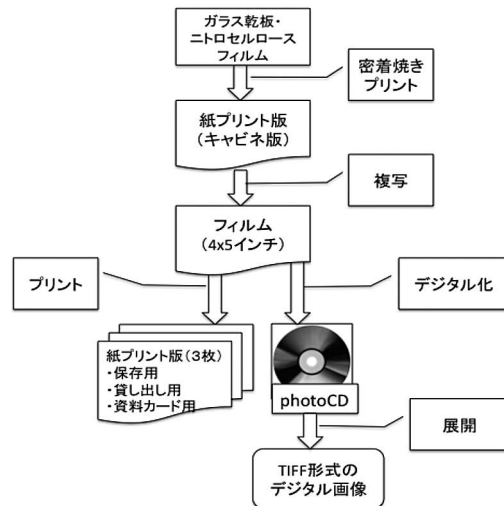


図4 歴博所蔵資料のデジタル化過程

まず、1992年にプリント版が作成されている。資料（ガラス乾板およびニトロセルロースフィルム）からキャビネ版用紙（120 mm×165 mm）へのプリント版が作成され、さらに4×5インチフィルム（105 mm×127 mm）への複写が行われた。元のプリント版への物理的接触を控えるため、この複写フィルムからのプリントが資料1点につき3枚ずつ作成された。その中の1枚は保存用として、元のプリント版とセットにして歴博の写真保管庫に収蔵され、他の1枚は貸し出し用として公開し、もう1枚は歴博内部者用の資料検索カードに貼付してある。図4は、この歴博所蔵資料のデジタル化過程を表したものである。

上記の過程に引き続き、2003年には複写フィルムをデジタルデータに変換してPhoto CD形式⁽¹⁾で記録した。このPhoto CDは、歴博の写真保管庫に保存されている。Photo CDに記録された可逆圧縮のオリジナル画像を非圧縮TIFF画像に展開し、歴博所蔵資料のデジタル画像データとした。

3.3 画像処理

取得されたデジタル画像に対する基本的な画像処理は、画像の回転処理、ホワイトバランスの調整、露出や明るさの補正などである。ホワイトバランスの調整は、光源にかかわらず白さが同じに見えるようにするように調整するものである。各処理は、デジタルカメラの事前設定機能を用いて行われたものと、デジタルカメラあるいはスキャナのそれぞれに用意されている画像取り込みソフトを用いて撮影後に行ったものがある。いずれの処理も、データの活用のために画像の見やすさを考慮し、必要な処理を施したものである。また、デジタルカメラ撮影に関しては、RAW画像で撮影し、それに対して画像補正を行うと同時にTIFFおよびJPEG画像へとファイル形式を変換した。

3.3.1 回転処理、ホワイトバランス、露出補正、明るさ補正など

画像取得の段階では、必ずしも文字や写真を見る上で適正な方向で取り込んではいないため、利

用の上で回転処理を施す必要のある画像については、画像処理ソフトのプレビュー画面上において目視で確認しながら1画像ずつ手動で行うか、コンピュータでスクリプトによる一括処理によって回転させるかという手段を適宜選択して行った。

事前のテスト撮影によって、資料を活用するうえで適正と思われるホワイトバランスや露出などを決定し、カメラの設定によって撮影時に適用する場合や、現像の際にその設定を画像処理ソフトにより反映させる場合があった。また、周囲の明るさや資料自体の色によってカメラの自動露出機構が働くため、必ずしも資料閲覧用として適正ではない露出となることがあった。また、資料によっては字の濃淡などが読みにくくなってしまう場合などがあったため、これらについては、現像後に一点一点目視によって確認した際に要再処理画像として選別し、画像処理ソフトによって個別に処理を適用した。

3.3.2 RAW 現像(画像フォーマットの変換を含む)

将来的に長く利用されることを期待して、汎用性の高い複数の画像フォーマットによってデジタル画像を保存することとした。汎用フォーマットとしてTIFFとJPEGを用い、これら2つの形式にRAW画像から変換した。この変換処理は、デジタル化作業の途中段階において、撮影画像と画像処理の結果を確認することにも役立てられている。ただし、作業効率やデジタル化作業時間の制約、画像のデータ容量などを考え合わせた上で、NMSの文書資料については、デジタル化作業の途中段階ではJPEG画像に変換することはなかった。既述のように、NMS所蔵資料は未整理の状態であり、資料名称や資料番号が確定していないため、活用を意図したJPEG画像は、本稿執筆時点でも作成されていない。

RAW画像については、画像変換における最上流データであることから、すべての資料について保存しておくこととした。万一、画像処理後のデータが失われてしまっても再現可能という意味で重要であること、異なるパラメータでの画像処理のやり直しが可能であるなどの機能上のメリットを重視したためである。

3.4 画像ファイルのリネーム

表7に記したように、デジタル画像のファイル名は、基本的に各所蔵機関によって与えられた資料番号を基準とするというネーミングルールに従ってファイル名をリネームした。このリネームによって画像ファイル名の重複を防ぐほか、ファイル名によって対象資料を知ることができるため、デジタル画像の管理と活用において利便性を提供することができる。

個別のネーミングルールとして、歴博資料の場合は「F-387-2-*-*-*」という資料番号のみであるため、そのままファイル名の基幹部分として使用している（*は0～9の数字。以下同様）。RAI資料では、「MS_249-*-*-*」の部分が資料番号であるが、同じ資料についてスキャンによる画像とデジタルカメラ撮影による画像の2種類が存在する場合があるため、スキャン画像であるのかデジタルカメラ画像であるのかを、SあるいはCという一文字を付すことで区別している。

開拓記念館資料では、「131 *-*-*」が資料番号であるため、これを画像ファイルの基幹部分として使用する。さらに、3.2.2節で述べたように、一つの資料につき3パターンの撮影画像があるため、

各パターンを区別するための識別子として透過光撮影には01を付し、フロント照明撮影のおもて面とうら面の画像には02あるいは03を付与することで区別している。

NMS資料の場合は、リネームは行っていない。その理由は、NMSによる資料整理が完了していないため資料番号が確定していないためである。したがって現段階では、文書資料（archive）と写真資料（photo）の区別をつけているのみで、デジタルカメラによ

って自動的に付される撮影順の4桁の番号のついた画像ファイル名のままである。しかし、それぞれの資料の内容と画像ファイル名との対応づけがなされるような資料情報の管理は行われている。

なお、表7のファイル名はTIFF画像の場合であり、JPEG画像の場合は拡張子がJPGとなる。

表7 所蔵機関とデジタル画像ファイル名

所蔵機関	ファイル名
歴博	F-387-2-*-*-.TIF
RAI	RAI_MS_249-*-*-.S*.TIF
	RAI_MS_249-*-*-.C*.TIF
開拓記念館	HMH_131*-*-.01.TIF
	HMH_131*-*-.02.TIF
	HMH_131*-*-.03.TIF
NMS	NMS_archive_*.tif
	NMS_photo_*.tif

3.5 作業中のバックアップ

本プロジェクト中に行われたデジタル化作業は、各々の期間において数日間にわたって続けられたため、作業中は基本的に作業日ごとのデータを、作業を行ったパソコン本体と外付けのハードディスクに一時的バックアップとして保存することとした。これによって、データ消失や誤った処理を施した上でのデータの上書き保存などの事態に備えた。したがって、一時的バックアップの方針としては、保存容量の許す限り、作業途中のすべての段階の画像データを保存しておくことにした。バックアップ作業は、デジタル化を行った時点での機材の環境によって異なるが、手動コピーまたはPC上のコマンドライン処理による自動化等の手法が用いられている。

また、所蔵機関におけるデジタル化作業が終了した後も、一連の画像処理作業は続けられたため、その過程でコピーされたデータや、画像処理の過程で生成された中間データも多く存在する。これらの中間データを保存しておくか否かは、画一的な判断を行うことはできないため個別に判断したが、その主たる判断基準は、RAW画像から何度でも同じ最終データを生成できるか、生成するための時間的コストはどれくらいかということである。

また、バックアップをとる際に特に気を使った点として、それぞれの画像ファイルや現像ソフトによって生成される画像補正パラメータファイルなどが含むタイムスタンプ、すなわちファイルの生成や変更の日時の情報の保存である。画像処理やその他の作業の段階においては、同じファイルネームのデータが複数生成されるが、それぞれのデータがいつの段階のものなのかを知る上で重要な情報であるため、上書き保存をしてしまったり、バックアップをとらずに編集してしまったりすることのないよう、十分に注意して作業を行った。

3.6 ネットワーク化による作業の分担

開拓記念館およびNMSでのデジタル化作業では、画像取得や画像の現像、バックアップなどの一連の作業を複数のパソコンを用いて役割分担させている。撮影と画像処理、画像の確認やバ

クアップをできるだけ同時進行させることで作業の効率性を高めるためである。各々のパソコンはデジタル化作業のために構築した極小規模ネットワークによって接続され、相互に画像データを利用できるようになっている。作業中に生じた問題の一つは、電力供給の不安定さに起因すると思われるネットワークの不安定さから、いくつかのデジタルデータにノイズが混入してしまったことである。このノイズは、画像の目視チェックによっても確認可能であるが、より万全を期するためにPCにおけるコマンドライン処理によって、コピー元の画像とコピー先の画像とをバイナリレベルで比較してノイズ混入がないことを確認している。このような危険もあるため、各段階でのバックアップデータの保存と、作業中の一つ一つの画像の確認が必要不可欠である。

4 デジタル化後の画像の保管および利用

4.1 保存用バックアップ

3.5 節で述べた作業中の一時的バックアップの他、本節では永続的に保存するデータのバックアップについて述べる。基本的な方針としては、撮影直後の RAW データとすべての処理を終えた TIFF 画像と JPEG 画像とを各機関資料について複数セットずつ保存することとした。

歴博所蔵資料については、複数のプリント版、フィルムおよび Photo CD の RAW 画像がある。TIFF 形式に変換してリネームしたデータは歴博に複数セット保存してある。RAI、開拓記念館、NMS の RAW 画像については、その資料の所蔵機関と歴博においてデータを保管している。表8は、RAW 画像と最終生成物としての JPEG 画像と TIFF 画像について、それぞれの機関がどのようなデータを所持しているかを示すものである。歴博には作業途中の中間生成データも未だ存在してい

表8 各所蔵機関が所持しているバックアップデータ

歴博	全機関の所蔵資料の RAW 画像
	全機関の所蔵資料の TIFF 画像・JPEG 画像
RAI	RAI 所蔵資料データの RAW 画像
	同リネーム後の TIFF 画像・JPEG 画像
開拓記念館	開拓記念館所蔵資料データの RAW 画像
	同リネーム後の TIFF 画像・JPEG 画像
NMS	NMS 所蔵資料データの RAW 画像
	同画像補正後の TIFF 画像

るが、歴博以外の各所蔵機関は、所蔵する資料についての RAW 画像および、すべての処理が済んだ TIFF および JPEG 画像をそれぞれ1セットずつ所持している。ただし、NMS は既述の通りリネームを行っておらず、文書資料の JPEG 画像は作成されていない。歴博では、それぞれについて複数セット所持している。

4.2 データベースの構築

本プロジェクトの全対象資料についてのデータベースを構築するため、デジタル化の作業と並行して、取得済みのデジタル画像に関する資料情報の入力を進めた。作成したデータベースの詳細事項を以下に記す。

4.2.1 資料情報データの項目と管理

それぞれの画像データに対応づけられるべき資料情報は、まず各所蔵機関ごとに Microsoft Excel（以下、エクセル）で管理している。そこには、表9に示す共通項目と、それぞれの機関に

固有の項目を設け、1 デジタル画像に関する一連の情報を一つのレコードとして入力した。各機関に固有な項目の例としては、RAI の画像回転処理に必要な情報や、スキャンの際に複数回に分割スキャンした場合、どの部分の画像にあたるのかを示す情報などがある。また、すべての機関のエクセルファイルを一つに統合したファイルを作成し、そこに各機関の個別ファイルに含まれている共通資料情報項目のみをコピーし、これをデータベースの原本として使用した。なお、本プロジェクトが日英間で進められているため、内容が日本語と英語で異なる項目については、英語の項目欄も用意している。

表9 資料情報データの共通項目

データの通し番号	公開／非公開	キーワード
所蔵機関	公開承諾先	固有名詞
資料番号	撮影年代	親戚関係
画像名	撮影地域	固有名詞に関する備考
画像取得年月日	資料概要	一致 *** 資料
資料の種類	撮影者または制作者	類似 *** 資料
材質	オリジナル／複製	関連 *** 資料
法量 (cm)	取得装置	
資料の物質的側面に関する備考	撮影に関する備考	

(*** は各所蔵機関名)

表9の項目について補足すると、「資料番号」はそれぞれの所蔵機関によってその資料に与えられたもの、「画像名」はその資料をデジタル化した後の画像ファイルの名前である。「資料の種類」は、「手紙」、「写真」、「スライド」、「箱」のように、資料の種別を記す項目であり、「材質」は「ガラス」、「紙」など、その資料の物質的な素材を記すものである。

さらに、一致資料・類似資料・関係資料の項目欄は、複数機関にわたって資料を比較できる機能をデータベースに持たせるためのものである。現状では主に写真資料に活用しているもので、一致するものは一致資料、類似性が見られるものは類似資料、関係のありそうなものを関係資料とし、相当する他機関の資料番号を一つのレコードごとに所定の欄に入力した。

4.2.2 データベースの構造

前節で提示した資料情報をエクセルで管理し、これをデータベースのための原本とした。この原本データをデータベース用のフォーマットに変換したうえで、データベースを構築した。なお、データベースの管理と運用には、オープンソースソフトを用いた。模式図を図5に示す。ただし、PCやソフトウェアではバージョンアップが頻繁に行われるため、恒久的にデータベースの管理と運営を行うためには、適切なタイミングでのデータの引っ越し、いわゆるマイグレーションが必須である。どのようなマイグレーション計画を立案し、確実に実行していくかはデジタルデータが有している根本的問題である。

データベースの利便性は、ユーザインタフェイスを含めた検索能力に左右される。そこで本プロジェクトでは、フリーワードでの検索と、それぞれの資料情報データ項目について検索項目を設定

できるような検索画面を設計して、ウェブアプリケーションとして実装した。データベースのトップ画面を兼ねた検索画面を図6に示す。検索でヒットした結果を一覧表示するページには、ヒットした画像データのサムネイル画像を表示する他、いくつかの資料情報項目についての情報内容も表示させた。そこから一つのデータを選択すると、選択された画像と情報が表示され、そのページには画像がより詳細に見られるように拡大機能も搭載した。さらに、その各画像ページの「一致資料」、「類似資料」、「関係資料」欄は、プルダウン形式で入力されている資料名を選択できるようになっており、資料名をクリックするとその関連資料にジャンプする機能も実装した。

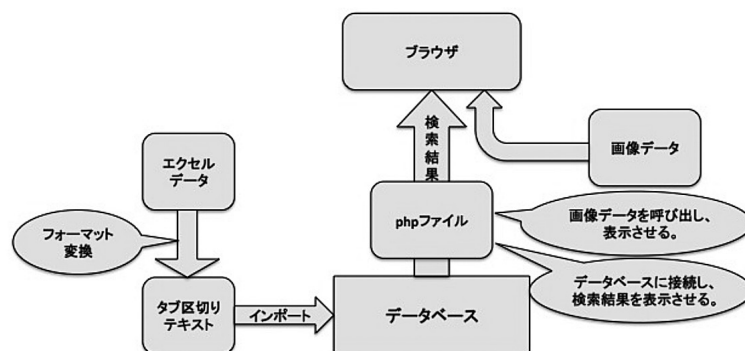


図5 データベースによる情報表示の流れ

マンロープロジェクト全部表示統合資料データベース

歴博、RAI、北海道開拓記念館、NMSのマンロー関係資料を検索できます。

Uchiage (alpha版) : ☒ AND検索 ☐ OR検索

検索語を入力したら、「検索実行」ボタンを押して下さい

複数の検索語をスペース（半角、全角ともに可）で分けて入力できます。
検索語の冒頭に半角の「-（ハイフン）」をつけると、その語を含まないデータを検索できます。

所蔵機関に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
資料番号に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
画像名に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
画像取得年月日に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
資料の種類に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
材質に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
資料複製に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
法量に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
資料の物質的側面に関する備考	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
公開に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
公開承諾先に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
撮影年代に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
撮影地域に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
撮影者または制作者に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
オリジナル性に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
取得装置に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
撮影に関する備考に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
キーワードに	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
固有名詞に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
親戚関係に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
固有名詞に関する備考に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
国立歴史民俗博物館の資料に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
王立人類学研究所の資料に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
北海道開拓記念館の資料に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない
国立スコットランド博物館の資料に	<input type="text"/>	を	<input type="radio"/> 含む	<input type="radio"/> 含まない

図6 データベースのトップ画面（検索画面）

5 今後の運用における課題

本プロジェクトにおける資料のデジタル化とは、対象資料のデジタル画像化だけでなく、その資料に関する情報もデジタル化し、さらにデータベースとして利用できるようになるまでの過程であり、具体的内容は前節までで記してきた通りである。本節では、その過程をふまえて、以下の5点について述べて、本稿のまとめとしたい。すなわち、デジタル化後のデータの保存ならびにその運用、資料所蔵者の権利関係に関わる課題、歴博所蔵資料についてデジタル化プロセスにおいて得られた知見、そして資料のデジタル化そのものが内包する問題、の5点である。

5.1 デジタルデータの保存

デジタルデータの保存の面では、「処理速度の点からデータ圧縮をしたほうが使いやすいと判断される場合でも、そうしてしまうとデータの保存形式が将来変わったときにその保存形式にあわせてデータ変換できるかどうかは保証されない。したがって、圧縮する場合でも圧縮前の状態を、保存用データとして位置づけておくべきである」ことがすでに指摘されている[笠羽, 2004, p79]。コンテンツのために利用するファイル形式はさまざまな画像処理を経て存在するが、少なくとも形式としてもっともシンプルな非圧縮のものを保存用として保持しておく必要がある他、将来においてもサポートされている形式である可能性を多く残しておくため、複数のフォーマット形式においてデータを保存しておく必要があると本プロジェクトにおいては判断された。

また、それとも関連して、デジタルデータがどれだけコピーを繰り返してもまったく劣化しないとか、何度加工を行っても同じ生成物が得られるという保証はない。本プロジェクトのデジタル化作業においても、撮影から画像処理やバックアップ作成の段階で、ネットワークを介して画像データをコピーするという過程が何度か繰り返されたが、コピーが正しく実行されたかを確認すると、デジタルノイズが生じていた例が少数ではあったが認められた。このような事態に備え、一つ一つの作業ごとに入念な確認作業も必要とされる。

5.2 デジタル化後の運用

文化資源をデジタル化することのみによって利益がもたらされるわけではなく、そのデジタルデータが効率よく利用できるかたちとなって初めてデジタル化の作業の意義が認められるものであろう。一旦デジタル化すると、データを適切に維持し続けていかなくてはならず、それにかかる人的・経済的コストも甚大である。デジタル化した後のデータの整備の仕方は計画段階から十分に吟味される必要があり、蓄積された資料の中から、利用したい資料をいかに効率よく取り出せるか、ということは、利用する人々の視点から、常により良い形を追求していかなくてはならない。

コンテンツのあり方をめぐっては、その構造や使い方がわかりにくくなることを避け、情報通信技術に接する機会が多くない利用者にも理解しやすい形態を常に指向していくことが求められよう。

5.3 データ所蔵者の権利関係

複数の所蔵機関、所蔵者による資料のデジタル化の場合、それぞれが守るべき法規が異なっ

るが、今回のデジタル化に際しては、肖像権、著作権や所有権なども異なる国の法規の下で判断する必要があった。デジタル化に際しては、法規の面で判断に慎重でなくてはならない、ということもすでに指摘されている通りである〔小川、2008 など〕。公開できるものか否か、できるとしたらどの範囲まで可能なのか、ということ、を、一点一点の資料について検討・確認し、それをコンテンツに反映させていくことも、地道に継続していかななくてはならない作業である。

5.4 歴博所蔵資料とデジタル化

このプロジェクトは歴博所蔵のマンロー資料を起点として広がったものであるが、デジタル化作業の中で明らかになった、他機関の資料のなかでの歴博所蔵資料の位置づけやその特徴をここで簡単にまとめておきたい。まず、歴博所蔵のガラス乾板資料およびニトロセルロースフィルム資料に写った内容には、プリント写真を撮影して作成されたと見られる複製ネガや、いくつかのオリジナル写真を組み合わせて1枚に仕上げた組み写真が複数点含まれている。これらの資料について、どこにあるどの資料が複製の元になったのか、さらにその元になった資料に付随する情報を引き出すことで、新たな知見が得られることもある。そのための複数の所蔵機関にわたる資料の比較が、デジタル化によって容易となったということがまず具体的な利点として挙げられる。

また、これは歴博所蔵資料に限ってではなく、資料一般について言えることではあるが、特に歴博資料はガラス乾板とニトロセルロースフィルムという非常にデリケートな資料であり、保存環境にも十分な注意を払わなくてはならないものである。そこに写っているものから読み取ることで、情報のみではなく、例えば資料そのものの物質的な情報など、実物を精査しなくてはわからないこともある。そのような場合に、デジタル化をすると同時に、物質的な情報も可能な限りデータとして書き留めておくことで、実物資料にあたって確認する頻度を減らすことになり、劣化を防止することにつながる。

5.5 「資料のデジタル化」が内包する問題

最後に、文化資源デジタル化という作業について、「アーカイブ」という概念に関連づけて若干の考察を行っておきたい。アーカイブをめぐる論争として存在していることの一つは、情報を一カ所に集約することがその占有によって権力性を帯びるなど、それが時に非常に権力の働く作業であるということである。そのようにならないよう、誰の利に与することになるのか、というソフトな側面は本プロジェクトのなかでも考慮する必要があった。法規や習慣などの違いからくる技術的、实际的、形式的な困難さに加え、それぞれの国や研究機関などによるその資料の位置づけや、利用の目的等も顧慮し、デジタル化作業そのものというよりも、データの保持の仕方や形態、アクセスの仕方などを互いに判断していかななくてはならない、ということもまた、難しい点である。

武邑光裕〔2003〕の指摘する通り、アーカイブとは、単なる記録／記憶ではなく、その生成には、後の世に何かを残そう、伝えようという、人々の「意識」が働きかけるものである。つまり、アーカイブ自体の持つ問題として、その情報選別に作為性が認められるということである。また、このプロジェクトでも常に大きな議題となってきた権利関係の問題にしても、その権利の所在を明確にしておくことをお互いに求め合ってきたのは、他者や他集団に関する情報を公開したり利用したり

すること、それ以前にその情報を所有しているということがいかにちからの働くことであるかを認めてのことであろう。

アーカイブがデジタルという形態をとることによってそこに拾われるもの／抜け落ちていくものがあることも指摘できる。例えば、本物の資料が持つ雰囲気や質感、においなどというものがある。しかし、そのような情報が必要とされない利用の範囲においては、それまで固有の文脈に位置づけられていた、または膨大な資料の渦のなかで埋もれていたそれぞれの資料、情報が互いに結びつき、別の文脈において新たな視座を獲得したり、意味の創造や変換を生じさせることには大きな意義がある。本プロジェクトを通してより明確に示された可能性は、その妙味を産み出す作業が、人工性・分散性によって特徴づけられるデジタルという形式によって、より活発に、より容易に行われることになるということだろう。

註

(1)——アメリカ Eastman Kodak 社とオランダ Philips 社との共同開発で Kodak の商標。拡張子は「.pdf」。
Portfolio II 版。

参考文献

小川千代子編 (2008) 『デジタル時代のアーカイブ』 岩田書院
笠羽晴夫 (2004) 『デジタルアーカイブの構築と運用』 水曜社
樫村雅章 (2010) 『貴重書デジタルアーカイブの実践技法—HUMI プロジェクトの実例に学ぶ』 慶應義塾大学出版会
慶應義塾大学デジタルアーカイヴ・リサーチセンター (2009) 『デジタルアーカイヴ—その継承と展開 慶應義塾大学デジタルアーカイヴ・リサーチセンター報告書 (2006-2009)』 慶應義塾大学出版会
後藤忠彦監修, 『デジタル・アーカイブ要覧』 編集委員会編 (2007) 『デジタル・アーカイブ要覧』 教育評論社
武邑光裕 (2003) 『記憶のゆくたて—デジタル・アーカイブの文化経済』 東京大学出版会
デジタルアーカイブ推進協議会 (JDAA) (2001) 『デジタルアーカイブ白書 2001』
研屋紀夫 (2009) 『デジタルアーカイブにおける「資料基盤」統合化モデルの研究』 勉誠出版

(お茶の水女子大学大学院人間文化研究科人間発達科学専攻,
国立歴史民俗博物館共同研究協力者)

(2010 年 9 月 27 日受付, 2011 年 2 月 21 日審査終了)