

【沖縄県公文書館】
【Okinawa Prefectural Archives】

Title	マイクロフィルムのデジタル化における高画質と低コストの実現：USCAR法務局文書の公開作業を通じて
Author(s)	堀川, 輝之
Citation	沖縄県公文書館研究紀要 = OKINAWA PREFECTURAL ARCHIVES BULLETIN OF STUDY(14): 33-56
Issue Date	2012-03-30
URL	http://okinawa-repo.lib.u-ryukyu.ac.jp/handle/okinawa/9141
Rights	沖縄県公文書館

沖縄地域学リポジトリ
ORION:Okinawa Repository Integrated Open-Access Network



マイクロフィルムのデジタル化における高画質と低コストの実現 USCAR 法務局文書の公開作業を通じて

堀川 輝之十

はじめに

- 1 公開を急げ
 - 1-1 上司からの指示
 - 1-2 公開の現実
- 2 外部委託せよ
 - 2-1 挫折
 - 2-2 餅は餅屋で
- 3 できるだけ実験を続けよ
 - 3-1 実験 1 手探り
 - 3-1-1 目的と方法
 - 3-1-2 結果
 - 3-2 実験 2 適度に
 - 3-2-1 目的と方法
 - 3-2-2 結果
 - 3-3 実験 3 両極端に
 - 3-3-1 目的と方法
 - 3-3-2 結果
 - 3-4 実験 4 クローズアップ
 - 3-4-1 目的と方法
 - 3-4-2 結果
 - 3-5 実験 5 ブレイクスルー
 - 3-5-1 目的と方法
 - 3-5-2 結果
 - 3-6 実験 6 合格
 - 3-6-1 目的と方法
 - 3-6-2 結果
 - 3-7 収穫

おわりに

はじめに

本稿では、USCAR 法務局文書¹の公開に伴う、デジタル化技術の確立について述べる。デジタル化の技術の確立とは、マイクロフィルムを「高画質で」「早く」「低価格で」PDFに変換する制作方法を作り上げることである。その技術を確立するために行なった実験の記録を紹介するのが、本稿の主な目的である。

本稿の独自性をより理解するために、筆者の紹介を簡単にしておく。まず、職歴について触れねばならない。筆者は、複数の職業を経験しているが、本稿で特に1つ挙げるべき職は、機械部品メーカーでのルート営業である。その営業の仕事は、ただ売り込みをかけて注文を取るだけではない。仕事の半分は、顧客の製品開発への協力や量産品のクレーム対応である。筆者は、それらの仕事の全プロセスに、コーディネイターとして、関わっていた。具体的には、技術者と議論を重ねながら、新規開発品およびクレーム対策品の両製作に関する事務作業を行っていた。その中で、筆者は、技術者の考える対策案が納期と価格の両面において現実的なものになるように意見の調整をしなければならなかった。このような仕事を通じて、筆者は、モノ作りの全工程を把握できるようになった。この経験が、今回の実験では、大いに役立った。品質向上と低コストを同時に実現させるために、スケジュール調整をしながら方策を練っていくという作業は、このような経験なしでは簡単にできるものではない。

次に、筆者の沖縄県公文書館での勤務について述べる。筆者は、当館で勤務するまでは民間企業での営利活動しか経験したことがなかったが、大学院で民俗学を学ぶ過程で、学術資料を扱う公的機関のサービスについて関心を持つようになった。

筆者は、当館で勤務する前は、公文書館の資料は工場の流れ作業のように容易に素早く処理されていると想像していた。筆者が勤めていたメーカーの工場では、コスト削減のために短時間で出荷できるような生産体制が整えられていた。生産ラインの各工程で無駄が省かれ、厳密な時間管理の下で、材料や部品、半製品が段取りよく準備され、加工され、組み立てられていた。しかも、絶えずさらなる工程の効率化が目指され、改善策が生み出されていた。そのために定期的に会合が開かれ、現場の知恵が総動員されていた。無理やりにもアイデアを出さないと許されない空気があり、その会合で指揮する各工程の班長は重い責任を担っていた。また、全員が正社員であり、現場を知り尽くしたベテランがいくらでもいたので、数十年単位での知識の集積があった。それが議論を深める力となった。

一方、文書館では、すべてにおいて工場のようにはいかないことがわかった。工場では効率よく生産を進めるのに適した材料や部品を選んで仕入れするのに対し、公文書館では、文書を受け入れる際に効率性を考慮に入れて選ぶということができないからである。たえず、不確実性の高い中で、膨大な数の多種多様な文書を処理していかなければならないという難しい状況に置かれている。また、嘱託員は最長3年までしか勤務できないので、その工場と比べると実作業の知識の集積ができていく環境にある。

それでも、筆者は、民間企業で得た経験、特に製造業で学んだことを生かしたいと思うようになった。製造業では「部材の仕入れから出荷まで」に様々な工程がある。一方、公文書館では「受け入れ

¹ USCAR とは、1950年12月から1972年5月まで沖縄を統治した、琉球政府の上位に位置した米国政府の出先機関。正式の名称は、琉球列島米国民政府 (United States Civil Administration of the Ryukyu Islands)。USCAR 文書とは、琉球列島米国民政府が作成・収受した約350万枚におよぶ資料群である。沖縄県公文書館は、平成9年度から15年度までの間、東京の国立国会図書館と共同で収集に取り組み、平成10年から順次公開している。本稿で扱う USCAR 法務局文書は、USCAR 文書全体の約30パーセントを占め、リール数にして562、コマ数にして1,056,026の分量をもっている。

から公開まで」において同じことが言える。その点で両者は類似している。類似しているならば、何らかの形で、製造業で得た知識を利用できるはずだと、筆者は考えるようになったのである。とはいえ、公文書の仕事については、赤子同然の素人。具体的に何に対してどのような形で貢献できるかは、すぐに思いつくことはなかった。

1 公開を急げ

1-1 上司からの指示

筆者は、2010年（平成22）英文チームの嘱託員として、主に、米国で収集された英文資料を整理・公開する作業を行ってきた。残念ながら、その年は、従来の作業方法を覚えるだけで精一杯であった。独自の工夫によって大きく貢献するところまでには至らなかった。2年目となる2011年（平成23）何らかの大きな成果を残さねばならないという焦りを感じるようになった。一通りの仕事を覚えた状態で前年と同じ質と量のパフォーマンスをしているようでは、進歩がないに等しいからである。さて、2年目は、直属の上司が変わることとなった。筆者は、同上司から、USCAR 法務局文書の公開作業に力を入れるよう指示された。その理由は、次の通りであった。

（1）2008年（平成20）までに法務局文書も含め全文書が公開済みだったはずだが、2011年（平成23）時点でまだ公開のための準備が完了していない²。新たな公開は2009年（平成21）以降中断している。

（2）国立国会図書館と共同で行なったプロジェクトゆえに、責任度が高い。

（3）USCAR 文書は、琉球政府文書および県文書と並んで、沖縄の歴史を理解する上で欠かすことのできない重要な文書である。特に、USCAR は琉球政府の上位に位置した組織だったので、本土復帰以前の沖縄の歴史を研究するには、USCAR 文書と琉球政府文書は対で参照されてこそより理解できる。

ちなみに、筆者は、この説明を受けた後に、もう一つ、USCAR 文書の公開を急ぐ意義を見出した。それは、USCAR 文書は、古文書ではなく、「新しい」ゆえに価値が高いということである。ここで言う「新しい」とは、USCAR が存在した時代の沖縄を知っている県民がまだ生存しているという意味である。「新しい」がために、同時代の目による検証が可能になる。つまり、USCAR の統治時代に生きた研究者や琉球政府に勤務していた人などが、USCAR 文書の記述が正確であるか、また補うべき情報が何か、言外にどのような意味が含まれているかなどを、現実的な感覚で調べられるのである。これにより、研究の確度が上がることが期待できる。

さて、この公開作業は、上述の理由の通り重要ではあるが、平成23年度の指定管理の事業計画には挙げられていなかったもので、事業計画で数値目標とともに挙げられた整理業務と同時並行で行わねばならなかった。そのような制約がある中で、高いパフォーマンスでこなすためにはどのようにすればいいのかを、筆者は考えた。

そこで、まず、筆者は、この重要な文書を、できるだけ多くの人を手軽に見られるようにするべきと考えるようになった。そのためには、ユビキタス（ubiquitous）化する必要があった。つまり、マイクロフィルムの資料をデジタル化して、いつでもどこでもオンライン上で閲覧できるようにする必要があった。同じUSCAR 文書の公安局文書では、デジタル化された資料（PDF）が公開済みである

² 沖縄県公文書館は2003年（平成15年）2月に、USCAR 文書の公開スケジュールを公表している。「一九九七年に始まったUSCAR 文書収集は、経済局資料の公開により、全体の約半分が閲覧可能となった。県は来年度に全資料の複写を終え、二〇〇八年に全面公開する予定だ」『沖縄タイムス』（2003年2月19日 朝刊）

が、現在は利用にあたって、制約がある。つまり、利用方法は、沖縄県公文書館内での閲覧もしくは有料での遠隔地複写サービスの利用³のどちらかしかない。筆者は、その不便な点を解消する方向で検討を始めた。そして、PDFをインターネット用のサーバーに載せて公開するという考えに至った。インターネット上の資料の閲覧やダウンロードが可能になるということは、県公文書館から遠く離れた場所で生活している県民（離島の住民など）との利便性格差が解消されるようになるということである。このような公平性のある税の使い方は、県の指定管理者である公文書館としては、考慮しなければならないことである。

その格差解消については、筆者個人だけでなく、沖縄県公文書館としても、重要視している。沖縄県が出す公募に対する、財団⁴側が提示する平成21年度の「沖縄県公文書館指定管理者指定申請書の提出について」の事業計画書には、遠隔地への情報発信が重視され⁵、そのための方法としてデジタル化が指定管理業務のひとつとして挙げられている⁶。また、第7回沖縄県公文書館に係る指定管理者制度運用委員会において発表された資料の中の「指定管理期間における重点項目」でも、「19年度～21年度」では全5項目の1つとして「ホームページの充実、デジタルアーカイブ機能の構築」、「22年度～24年度」では全6項目の1つとして「デジタルアーカイブ事業の充実強化」が記載されている。

また、実際に、2011年（平成23）11月に寄せられたあるアンケートで、利用者から次のような声も出ている。「宮古島から初めて入館いたしました。公文書館準備段階から貴館のことは知ってはおりましたが、機会がなかなかつくれませんでした。今後大いに利用させていただきたいと思います。遠隔地の者のためにインターネットによる公開をスピードアップかつ充実させていただきたい。国会図書館の方が身近に感じる現状です」（60代）。このような声に注意深く耳を傾けていくべきであろう。

1-2 公開の現実

これまでの収集と公開の状況は図1のようになっている。収集作業は平成15年度に終了している。公開については、平成14年度に、USCAR文書全体を2008年（平成20）までに公開すると発表したが、現時点で、百万コマ以上からなる法務局文書の公開がまだ残っている。「2008年での全公開を想定した累積コマ数の推移」と「実際の累積コマ数の推移」については、図2で比較して見ることができる。また、図1を見ると、「H13年度」から公開が隔年となり、「H21年度」以降は公開が中断されていることがわかる。

³ データをCDなどの媒体に入れた形で利用者に郵送するサービス。

⁴ 公益財団法人沖縄県文化振興会。

⁵ 「沖縄県公文書館は地方自治法第244条に規定する「公の施設」として、県民の利用について不当な差別的扱いを行わないことを基本とする。（略）県内離島をはじめとする遠隔地居住者への情報発信を重視し、来館困難な県民へのサービス提供の拡大に取り組む」（pp.3-4）。

⁶ 「利用頻度の高い公文書等をデジタル化しウェブで閲覧に供する」（p.53）。

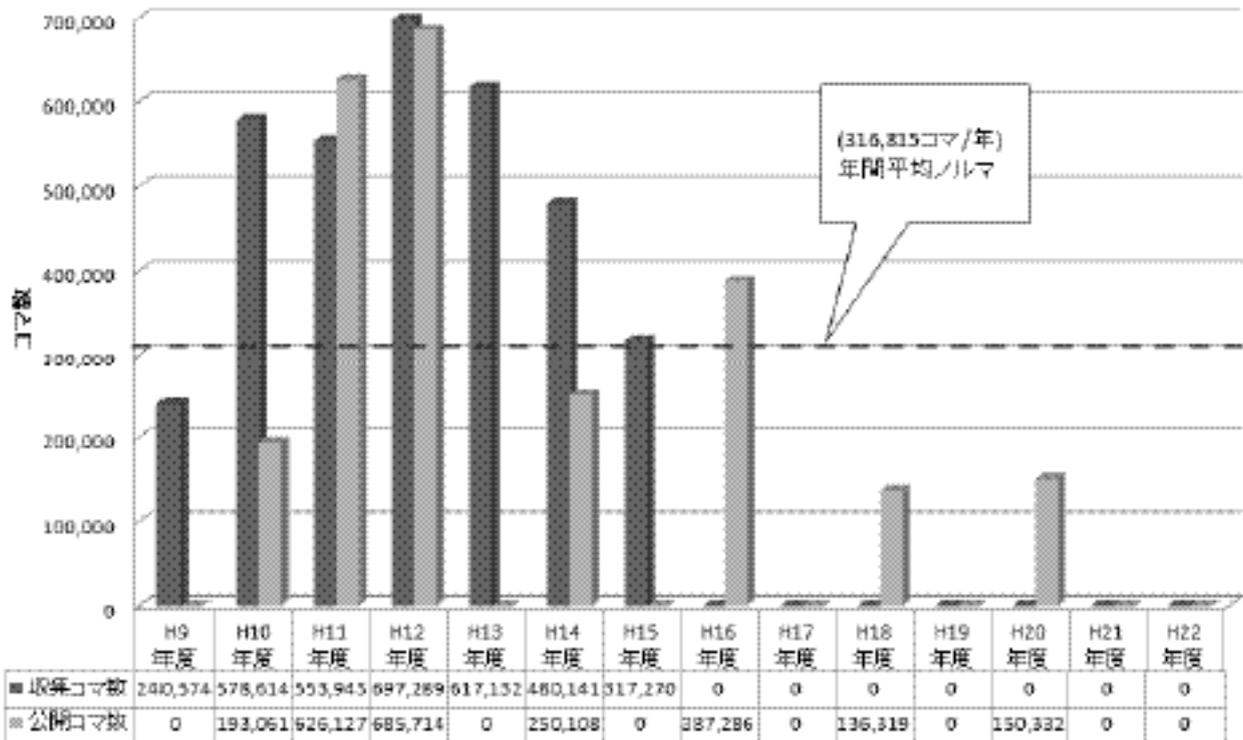


図1： USCAR 文書における収集コマ数と公開コマ数の推移

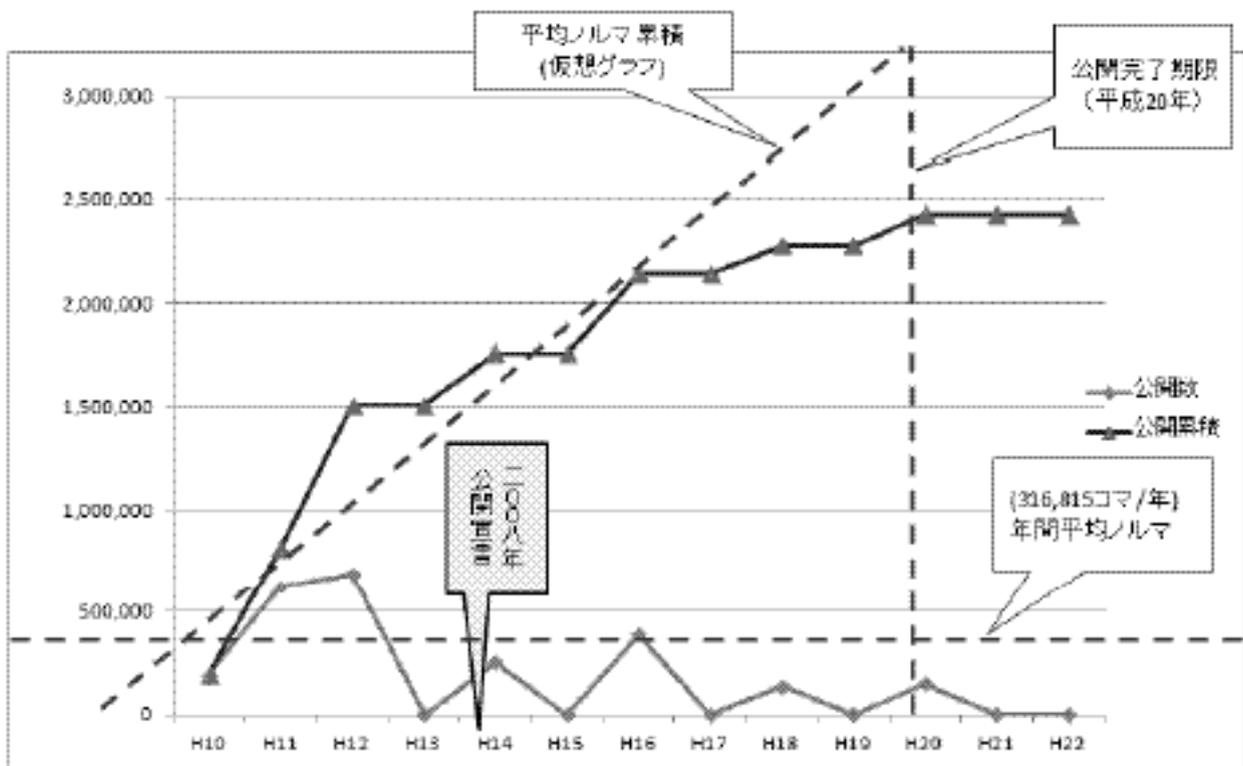


図2： USCAR 文書における公開コマ数の推移（各年度数および累積数）

2 外部委託せよ

2-1 挫折

前年の後半、筆者は、USCAR 法務局文書のマイクロフィルムをデジタル化する計画を立て始め、まずは、当時すでに当館にあったマイクロスキャナー MS6000 (Konica Minolta 製) を利用しようと考えた。筆者は、前年の上司からも、USCAR 法務局文書の公開作業に着手する計画について聞いていた。しかし、2011年(平成23)になって、筆者は、具体的な制作スケジュールを作成しながら、同スキャナーの性能を確認していくと、処理速度と画質の2点において問題があるということがわかった。

まず、処理速度については、実機で性能を確認することにより、計画で求められる基準からすると不十分という結論を出すことができた。法務局文書のマイクロフィルムには、1リールにつき1879のコマが収められている⁷が、解像度をMS6000では最も低い400dpiに設定しても、1リールのデジタル化に6時間半から7時間の時間を要することがわかった。これでは、1日1リール分しか処理できない。その処理速度で、同文書の全リール562本分をスキャンすると、約25ヶ月間かかる。2年以上である。これでは、とても急ぎの公開とはならない。2008年までに全資料を公開すると公表したにもかかわらず、2011年にまだそれが完遂できていないどころか約30パーセントのボリュームも残しているという状況を考えると、スキャンだけで2年以上かかるのは許容できないと判断せざるをえなかった。

次に、画質面について述べると、元画像(マイクロフィルムのコマがモニターに映し出されているときの画像)の画質が悪い場合、スキャン後、「きわめて読みにくい画像(判読できる文字が半分以下の画像)」もしくは「まったく読めない画像」になってしまうということがわかった。たとえば、カーボン紙を使ってタイプ打ちしたと思われる資料では、文字以外の余白の部分が黒ずんでいる(図3を参照)。この部分が、少し濃くスキャンされるだけで、文字との判別が難しくなるほど濃くなってしまふ。その他、印字に使われたインクが悪かったためか、かすれたり、滲んだりしている文字もある。これらの文字を精細に読み取ってデジタル画像に変換するのは難しい。筆者は、MS6000で調節可能なモードである解像度と濃度、コントラストの数値を様々なパターンで組み合わせて実験をしたが、芳しい結果を得ることはできなかった。

結論として、同文書のデジタル化に関していえば、MS6000は、処理速度とスキャン性能の2点からみて、利用は難しいということになった。その結果、デジタル化の外部委託を検討することとなった。

2-2 餅は餅屋で

筆者は、MS6000の納入・メンテ業者と別件で打合せ中に、USCAR 法務局文書のデジタル化の問題点について触れた。同社はUSCAR文書のマイクロフィルムの製作を担当し尚且つデジタル化業務において当館への納入実績を持つ業者なので、筆者は同社から何か問題解決のヒントを得ることができるかもしれないと考えたのである。すると、同社が所有するマイクロスキャナー RS200 (ScanStation 製) で、高速スキャンができるということがわかった。1リール分の処理速度は、

⁷ コマ総数が1,056,026で、マイクロフィルムの総リール数が562。よって、次の計算式により、マイクロフィルム1リールの平均コマ数は、1879となる。 $1,056,026 \div 562 = 1879$ 。

⁸ RS200とMS6000の比較をした時のPC環境はそれぞれ次の通りである(CPU、搭載メモリ、OSの順で列挙する)。RS200: Intel CORE2 DUO、4GB、Windows 7。MS6000: Intel COREi7、4GB、Windows 7。

机上の大まかな計算で約1時間、つまりMS6000の約7倍である⁸。筆者は、これは検討を進めていくに値する速度と判断し、早速、実機を使ってその速さが実現できるかどうかを同社に確認してもらうことにした。

当館は、12本のマイクロフィルム⁹を同社に貸し出し、同社は、それらを使って、制作にどの程度の時間を要するかを算出した。その結果、スキヤニングの速度は1リールにつき約30分ということがわかった。これは、MS6000で要する6時間半から7時間と比べると13~14倍も速くスキヤニングできることを意味する。ただし、それで作業時間を13~14倍削減できるようになるというわけにはいかなかった。なぜなら、スキヤナーが区切る箇所を誤って画像を取り込むことにより、そのコマの情報が1つの画像におさまらない場合も出てくるので、結局2回スキヤニングをして、「誤って読み取られた画像」を「正しく読み取られた画像」に差し替える工程も入れなければならないということがわかったからである。つまり、2回のスキヤニングにより1リールにつき60分間かけ、尚且つ「読み取りの区切りが正しい画像」を選ぶ工程も追加しなければならなくなったのである¹⁰。制作全体を考えると、その他、品質保証の検査や納入媒体(DVD-RW)への記録など様々な工程に要する時間も加算しなければならない。それらを全て含めると、机上計算ではあるが、約2時間で1リールを作成できるという結果になった。これをもとに計算すると、7ヶ月強で562リールの制作が可能となる(1日4リール、1ヶ月80のペース)。MS6000の25ヶ月と比較すると4倍以上の早さで制作できるようになる。明らかに処理速度に差があるので、外部委託の方向で検討を継続することとなった。餅は餅屋でという考え方に切り替えたのである。次の課題は画質であった。

3 できるまで実験を続けよ

本節では、画質を向上させるために、いかなる実験をしてきたかを、時系列で説明する。実験は、2011年6月初めから8月末にかけて、大きく6回行われた。全ての実験で、同社所有のRS200が使われ、同社の社員の方に同スキヤナーの操作をしていただいた。実験の後は、毎回打合せをし、問題点を明確にしながらか解決策を練っていった。以下はその軌跡である。

3-1 実験1 手探り

3-1-1 目的と方法

実験1では、どのようなタイプの画像がスキヤニング後に「読みにくくなるか」または「読めなくなるか」を明らかにすることにした。そうやって不合格になる傾向をつかむことができれば、その後の実験では適切な対策を講じていけるようになると思ったのである。

さて、実験は上述の12リールで行われた。全てのリールにおいて、濃く写っているコマと薄く写っているコマが共に数多く含まれているということに気づいた同社は、様々な濃淡に対応する方法を考えてスキヤニングすることにした。具体的には、次の手順でテストすることにした。1) マイクロフィルムの濃度を、3か所(頭の方、中間、後ろの方)計測し、それらの平均値を出す。2) その平均値に対応するプロファイルを作成する。プロファイルとは、トラッキング小、トラッキング大、トラッ

⁹ これら12リールは、全562あるリール(リール番号1-001から61-013まで)の中から、次の通り、万遍なく選んだ。1-001、5-006、10-003、15-004、20-003、25-008、31-004、36-007、41-005、46-003、52-008、57-004。

¹⁰ この問題は、実験を重ねていくうちに、頻繁に起こるものではないということがわかったため、この問題の対処のためにわざわざ毎回2回スキヤニングする必要はないということになった。

¹¹ トラッキング小：フィルム画像の濃さの下限を決める、トラッキング大：フィルム画像の濃さの上限を決める、トラッキング境界：黒枠などの枠の境界を決める、しきい値データ：テキストなどの情報の輝度値のしきい値を決める、

キング境界、しきい値データ、しきい値背景、明るさ、コントラスト、ランプ照度、フィルタ大きさ、フィルタ感度¹¹における各設定値の組合せのことである。3) 2) で作ったプロファイルで実際にスキヤニングする。

3-1-2 結果

12のルールでスキヤニングして作成したPDFファイルの不合格率は、同社独自の判断では、各々の通りとなった。49.1%、14.4%、9.0%、3.9%、10.9%、5.3%、16.3%、1.0%、18.2%、4.8%、8.0%、5.0%。平均12%である。筆者は、スキヤナーの能力に限界があるということを差し引いても、この不合格率は高すぎると考えた。基本的に、不合格率はゼロでなければならない。よって、画像の文字の判読性がスキヤニング後に低下する原因を突き止め、それに対する策を講じていかねばならない。筆者と同社担当者は、元の画像が濃すぎたりまたは薄すぎたりすると不合格画像ができやすいと

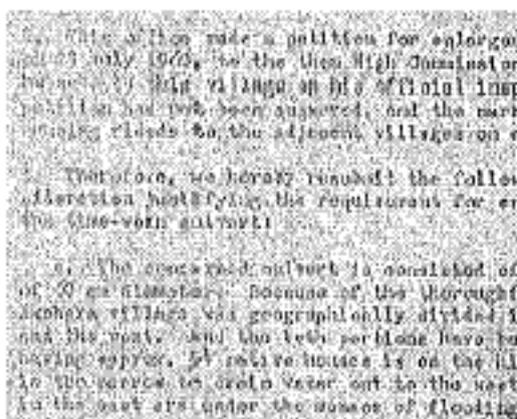


図3：不合格になりやすい濃い画像

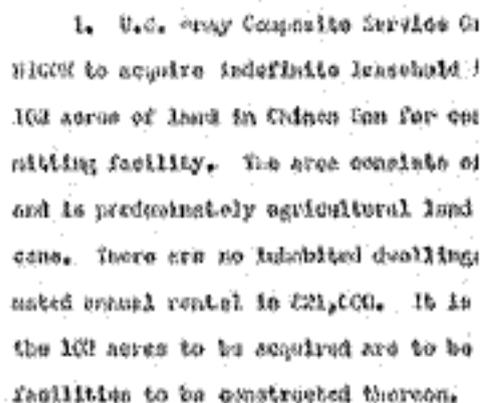


図4：不合格になりやすい薄い画像

判断し（図3および図4を参照）さらに細かく濃淡に対応できるようなプロファイルを作成する必要があるという結論に至った。

ちなみに、実験1では、各画像は、同社だけの主観的判断によって不合格と合格に分けられた。しかも、厳格な基準をもって判断されたのではなかった。「何となく画像が悪い」「読みにくい部分がある」「かすれている文字がある」というあいまいな基準で合格と不合格の判別が行われたのであった。筆者と同社担当者は、その点を見直し、以降の実験では、明確な判断基準に基づき不合格か合格かを決めなければならないということを確認した。そこで、次のような品質基準を設けることになった。

(1) スタンプ（図5を参照）や手書き（図6を参照）の情報は、元のマイクロフィルムの画像でも、すべての文字が読める場合がきわめて少ないので、判読できなくても仕方がないとする。つまり、判断の対象から外す。タイプ打ちされた文字が読めるかどうかを重要視することにする。

しきい値背景：背景（余白）の輝度値のしきい値を決める、明るさ：背景（余白）の濃さを決める、コントラスト：グレースケールかハーフトーンかを決定し、ハーフトーンモードのときは濃い色調にする、ランプ照度：ランプの明暗度を決める、フィルタ大きさ：画素行列の差により2次元フィルタ機能の効果の大きさを決める、フィルタ感度：2次元フィルタの感度を決める。

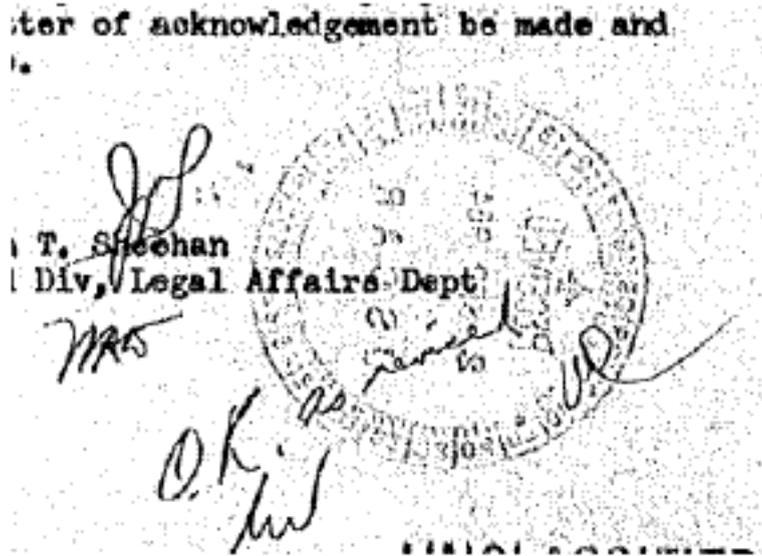


図5：文字がかすれていて判読できないスタンプ

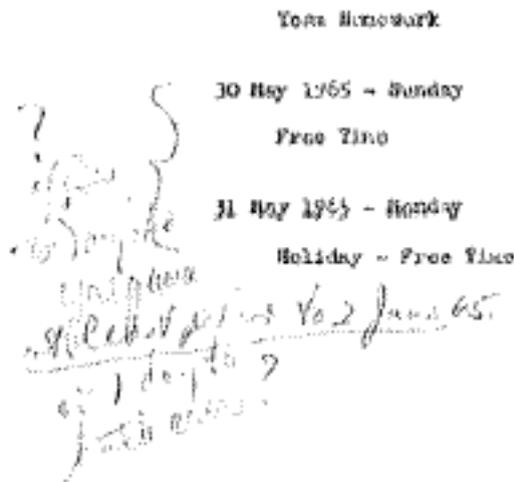


図6：文字がかすれていて判読できない手書き文字

- (2) 手書きだけのページは、おおよそ再現できていれば合格とする。
- (3) モニターに映しだされたマイクロフィルムの画像（マイクロスキャナーでスキャンする前の画像）にある文字が大部分判読できない場合、スキャン後の画像がいくら悪くても不合格にしない。つまり、元の画像が読めないものはどうしようもないと判断する。
- (4) 文字数が多い画像では、読めない単語が3~5つ程度混在していても、判定は合格とする。スキャンする前の時点ですでに、一部分だけ文字がかすれたりもしくは潰れたりして読みにくいという画像は多いので、この程度の難読性は許容する。

3-2 実験2 適度に

3-2-1 目的と方法

実験1の結果、様々な濃さに対応できるスキャンが求められるようになった。それに応えるた

めに、同社は、15のプロファイルを作成した。元画像が薄目にできている場合、濃く仕上げるプロファイルを使う。逆に、元画像が濃い目の場合は、薄く仕上げるプロファイルを使う。ところが、「濃い」として「薄い」は、それぞれ1種類ではない。よって、プロファイルは「濃い」と「薄い」の2種類のみというわけにはいかない。様々な濃淡の画像に対して、それぞれ適切なプロファイルを用いてスキャンしなければならない。そのために、15種類ものプロファイルが用意されたのである。

実験2の具体的な手順は、次の通りとなった。1)実験1同様、マイクロフィルムの濃度を、3ヶ所(頭の方、中間、後ろの方)計測し、それらの平均値を出す。2)1)で出た濃度に適していると思われるプロファイルを「15のプロファイル」から2つ選び、それらを1回ずつ用いてスキャンする。3)スキャン後、2つのうち、良いほうの画像を選ぶ。

ちなみに、本実験では、上述の12のリールから無作為に3つのリールを選んで用いた。

3-2-2 結果

実験の結果は次の通りである。

(1) 全体的に濃く見える元画像、つまり文字以外の部分(余白)に「黒粒」がたくさん入ったような状態になっている元画像は、スキャン後、「黒粒」の少ない画像になった。これにより、文字の部分が見やすくなった。本実験の方法は、元画像が濃い場合には効果があるという結論がでた。

(2) 全体的に薄く見える元画像(図3参考)、つまり文字がかろうじて読める状態になっている元画像(図4参考)は、スキャン後、文字の見えないもしくは見えにくい画像となった。本実験の方法は、元画像が薄い場合には、画質の低下を招くという結果がでた。文字の薄い画像が多く入っているリールの場合には、不合格率は高くなると予想できる。よって、この点を改善する必要がでてきた。

3-3 実験3 両極端に

3-3-1 目的と方法

実験2の方法では、濃い元画像に対しては良い結果が出たが、薄い元画像に対しては逆に悪い結果が出た。実験3では、その差をなくすことが求められた。つまり、薄い元画像と濃い元画像の両方において、スキャン後も文字が読めるようにしなければならないということである。筆者と同社は対策を考えた。まず、実験3では、実験2と同様、1リールにつき2回スキャンするという前提で考えることにした。そうすると、1回目と2回目のそれぞれのスキャン方法をどのようにするかという点が問題となった。

1回目については、実験2で問題となった薄い画像に対応する方法を検討した。そして、不合格数を極力減らすためには、きわめて薄い画像の再現性も高める必要があると考えた。その結果、「薄い元画像は濃くスキャンすればいい」という考えに行きついた。しかも、「きわめて薄い元画像へも対応できるように、かなり濃くスキャンすればよい」ということになった。

ところが、「かなり濃くスキャンする」ことにより、デメリットが出ることも予想できた。つまり、元々濃い画像を不必要に濃くスキャンすることにより、文字が潰れて読みにくくなったり、文字以外の部分(余白)の「黒粒」が濃くなったりまたその数が多くなったりする現象が起こると推測できたのである。もし、濃い画像が多く含まれているリールをスキャンした場合、この方法で生じるデメリットは大きくなるであろう。

そこで、筆者は、1回目は「かなり濃くスキャン」、2回目は「かなり薄くスキャン」する

という方法を思いついた。こうすれば、かなり濃い元画像へもかなり薄い元画像へも対応できると思ったのである。しかも、この「濃淡両極端方式」では、両プロファイルのデメリットを相互に打ち消す効果もあると想像できた。

とはいえ、両極端なプロファイルを使うということは、「やや濃い」または「やや薄い」元画像の画質を犠牲にするということを意味する。つまり、15のプロファイルを用いて得られるメリットを放棄することになるのである。筆者は、そのリスクは仕方がないと考えた。「やや濃い」「やや薄い」などの画像は、濃淡の極端なプロファイルでスキャンされた場合、多少読みにくくはなっても、読めなくなるということはないであろうと考えたのである。要するに、「なんとか文字が読めればいい。最終的に不合格率さえ下がればいい」という発想で、前の実験で明らかになった課題を克服しようと思ったのである。

まとめると、実験3の方法は次のようになる。1)1回目、かなり濃くスキャンする。2)2回目、かなり薄くスキャンする。3)比較して良いほうの画像を選ぶ。

ちなみに、本実験で用いたリールは、濃い画像が多く含まれているもの1つと、薄い画像が多く含まれているもの1つである。濃淡面で条件が厳しいリールを用いることで、「濃淡両極端方式」の効果が見えてくるであろうと考えたのである。

3-3-2 結果

実験の結果は次の通りである。

(1) 不合格率は5%以下になった。

(2) 「かなり薄い元画像」は、かなり濃くスキャンされた後、かろうじて読める水準で再現されていた。一方「かなり濃い元画像」は、かなり薄くスキャンされた後、特に目立って文字が薄くなるということはなく、見やすいものになっていた。

(3) 「やや薄い元画像」と「やや濃い元画像」については、かなり濃くスキャンするほうが、かなり薄くスキャンするよりも、不合格になりにくいということがわかった。よって、かなり薄くスキャンする利点は、「かなり濃い元画像」を文字潰れなしに再現できるということのみということがわかった。

(4) 確実に不合格になる画像は、次のような種類であることがわかった。「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる（水が多く含まれる、磨り足りない墨で半紙に字を書いたときのように見える）」の3種類である。たいていの場合、これらのうち2種類以上が組み合わせられて文字が読みにくくなっている。文字だけを確実に選び出すというスキャンができていないのである。言い換えれば、文字と文字以外の部分の区別が中途半端なままスキャンされているということである。よって、薄くスキャンされた場合、元画像ではかろうじて可視できる「文字の芯」（滲んだ部分を除いた、本来あるべき線）が、スキャン後には、部分的に薄れたり、消えたりするのである。また、濃くスキャンされた場合、「滲んで広がった部分（本来不必要な部分）」を「文字の芯」と誤って認識してしまうという現象も起こっている。つまり、文字潰れの現象である。例えば、a や c、o、e、d、s、g などの形が似ているアルファベット文字がすべて同じように見えるようになり、区別できなくなっている。これらの文字は、x や z などとは違って、単語を構成する頻出文字なので、多くの単語の判読ができなくなるという結果を招く。

(5) 合格と判断された画像であっても、(4)で述べた「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」のうちのどれかに当てはまる場合には、スキャン後には、元画像と比べて極端に画

質が落ちるということがわかった。

(6) 実験3の方法では、2回のスキャニング後により方の画像を選ぶという作業が必要不可欠であるが、この選択作業には、2つの大きな問題があることがわかった。1つは、選択が容易にできないということである。品質基準を正しく理解して、多種多様な画像を合格か不合格に区別するという作業は、かなりの慣れが必要である。また、ほとんど全部の資料が英文なので、「読めるか読めないか」の判断は日本語の場合よりも難しい。もう1つは、莫大な時間を要するという点である。1リールにつき約2000コマを目視して選別を行わねばならない。1コマ1秒で見た場合33分間要する。これは、判断に悩む画像がまったくない場合である。それでも30分以上かかるのである。実際には、10数秒間考えさせられる、判断の難しい画像が数百という数で入っている。かりに10秒間考えねばならないコマが1リールに300あったとすると、3000秒(50分)要することになる。大雑把にみても、この選別作業だけに、1時間はかかると推測できる。実作業を行った担当者も、「この選別作業に数時間を要する」と言っている。これでは、高速マイクロスキャナーを利用するメリットがなくなってしまう。冒頭でも述べた通り、デジタル化の技術の確立とは、「高画質で」「早く」「低価格で」PDFに変換する制作工程を作り上げるということである。「早く」制作することには2つの重要な意味がある。1つは、すでに全面公開の時期が過ぎている状況でできるだけ早く公開するという責務を果たすことができるということ。もう1つは、工程の時間短縮により、人件費を抑えることができるという点である。「早く」は、「低価格で」に連動している点である。よって、時間短縮のための方法を見出すことは重要で、何とかして、工程の簡素化・時間短縮化を考えねばならなくなった。

3-4 実験4 クローズアップ

3-4-1 目的と方法

実験3では、画質と時間の2点に関する問題が明らかになった。まず、画質についてであるが、実験3の方法では、「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」という特徴を持つ画像を高い精度で再現することはできないということがわかった。そこで、筆者は、「もしこれらの特徴を持つ画像が1つのリールに多く含まれていたら、不合格率は激増するのではないだろうか」と考えた。筆者は、不安になって幾つかのマイクロフィルムを調べてみた。すると、それら3つの特徴をもつ画像が多く含まれるリールが見つかった。しかも、調べれば調べるほど、これらの種類の画像は、珍しくないどころか、標準的と言っていいほど多く存在することがわかった。画質の向上に対する考え方を根本的に変えなければならなくなった。それまでは、「幅広い濃淡のレンジがある中でどのように濃さを合わせていけるか」という考え方に基づき対策を打ってきた。しかし、それは誤ったアプローチということがわかった。濃淡調節によるスキャニングには限界があるということが明らかになったのである。そこで、筆者は、「濃淡」という考え方を一度完全に捨てて、新たな視点から考えねばならないと思うようになった。

さて、新たに考えついた見解は、次の通りである。もっとも大事なことは、「正確に文字の芯だけを読み取る」ということ、つまり「文字と文字以外の情報を明確に区別する」ということである。それさえ実現できれば、濃淡など関係なしに「読める画像」を作ることができるはずである。文字が小さかったり、薄かったり、滲んでいたとしても、文字の芯だけをしっかりと捕えることができれば問題ないのである。その結論に至った筆者は、具体的にどのようなスキャニング方法によってそれを実現できるかを検討した。

実験4では、次の方法をとることになった。1)「文字サイズが小さい」、「文字が薄い」、「文字が滲

んでいる」という特徴が全部揃っていて、なおかつ実験3で不合格となった画像を1つ選ぶ。実験4では、この画像だけを用いる。再現のきわめて難しいこの画像で画質の向上が認められたら、実験成功と判断できる。2)倍率を上げてスキャニングする(小さな文字は大きくしたら見やすくなるだろうという単純な発想に基づく策である)。具体的には、27倍(標準の倍率)より高い倍率(29.7倍、32.4倍、35.1倍、37.8倍、40.5倍、43.2倍、45.9倍、45.9倍、47倍、48.6倍、50倍)でスキャニングしてみて、もっとも効果がありそうな倍率を1つ決める。3)「文字と文字以外の情報を明確に区別する」プロファイルを作成する。それまでの濃淡調節型のプロファイルとはまったく異なるものを作らねばならないので、多くのプロファイルを試してみる必要がある。トラッキング小、トラッキング大、トラッキング境界、しきい値データ、しきい値背景、明るさ、コントラスト、ランプ照度、フィルタ大きさ、フィルタ感度の各数値を変えながら、様々な組合せを作るのである。明らかに意味のない組合せを除外したとしても、2万4千以上の数の組合せがある。もちろん、その数のプロファイルをテストする時間的余裕はなかったので、同社が566まで絞り、最終的には筆者が216まで絞り込んだ。4)2)で選んだ倍率でもって、この216の組合せをすべて試す。

ところで、もう1つの問題である時間についてであるが、これは、画質の問題と連動している。実験3では、「かなり濃くスキャニング」だけでは、画質の問題を解決しきれなかったので、「かなり薄くスキャニング」も加えることになった。そして、これらのプロファイルで作った2つの画像から良いほうを選ぶという工程も加えることになった。しかし、かりに、どのような画像にも対応できるプロファイルで、正確に文字の芯を読み取ることができるようになれば、スキャニングは1回で済むようになる。その結果、自動的に1回目と2回目の画像の選別も行わなくていいようになる。これにより、「選別に要する約60分間」と「スキャンニング1回分に要する30分間」の計90分間が浮くようになる。すなわち、画質の向上が可能になることで時間の問題は解消されるようになるということである。

3-4-2 結果

実験の結果は次の通りである。

(1) 図7の通り、40.5倍の倍率でスキャニングすると、標準倍率の27倍(図8を参照)と比べて明らかに見た目の印象が変わった。文字全体がやや引き締まったように見えるようになったのである。40.5倍より低い倍率では、標準設定の27倍の場合と比べて、見た目上の変化は見られなかった。また、40.5倍より高い倍率では、40.5倍との違いを見つけることができなかった。40.5倍で見られた変化が「正確に文字の芯を読み取る」ことにつながっているかどうかはわからなかったが、40.5倍で216のプロファイルをテストすることにした。

1. The Naval Air Facility, Faha, advised this is the following range will be used by the U. S. Navy:

Sekibi Sho (Also known as Raleigh Rock, Akao

From 0900 hours to 2100 hours, 22 Sept 1959.

2. Request your office publish notice of same to

3. All departments which have been furnished a requested to advise their counterpart of ONI to disseminate to all concerned.

4. A radio Message will be dispatched to the re YCAT for their information.

図7：40.5倍でスキャンした画像

1. The Naval Air Facility, Faha, advised this is the following range will be used by the U. S. Navy:

Sekibi Sho (Also known as Raleigh Rock, Akao

From 0900 hours to 2100 hours, 22 Sept 1959.

2. Request your office publish notice of same to

3. All departments which have been furnished a requested to advise their counterpart of ONI to disseminate to all concerned.

4. A radio Message will be dispatched to the re YCAT for their information.

図8：27倍（標準倍率）でスキャンした画像

(2) 40.5倍で216種類のプロファイル进行测试した結果、同社は9つのプロファイルが比較的よいという結論を出した。筆者は、その9つのプロファイルでできたPDF画像を詳細に見て、さらに絞り込んでいった。画像の優劣をつけるために、「aやc、o、e、d、s、gが潰れずに原型をとどめているかどうか」を見るという方法をとった。最終的に、2つのプロファイルを選びだした。それらは、「組合せ番号10」および「組合せ番号19」と称していたものであった。

(3) (2)で選んだ2種類のプロファイルで作った画像（40.5倍でスキャン）と標準設定の27倍でかなり濃くスキャンした画像を比較してみると、「文字の見やすさ」という点において、それら3種類にほとんど差は見られなかった。27倍で作られた画像のほうが見やすい箇所もあった。これに

より、「倍率を上げ」なおかつ「厳選されたプロファイルを使って」も、「文字と文字以外の情報を明確に区別する」という効果を出すことはできないということがわかった。実験失敗である。

(4) 40.5倍でスキャンすると、27倍でスキャンしたときの2倍から3倍の時間を要するということがわかった。これは、40.5倍をあきらめるダメ押しの原因となった。

3-5 実験5 ブレイクスルー

3-5-1 目的と方法

実験4で、「倍率を上げ」なおかつ「厳選されたプロファイルを使って」も、良好な結果をえることはできなかった。これ以上打つ手はないと思われたが、筆者は、あきらめずに他の手を考えてみた。思いついたのは、倍率を上げることによってむしろ悪影響が出たのではないだろうか、ということである。そこで、こう考えた。倍率を元に戻し、その上で216種類から厳選した2つのプロファイル(「組合せ番号10」と「組合せ番号19」)を使えば、よい結果をえられるのではないだろうか、と。その方法しか残されていないと思えた。それしかないのならそれを試すしかない。筆者は、最後の望みを託すように、倍率を標準の27倍にして、「組合せ番号10」と「組合せ番号19」でテストするよう同社に依頼した。テストで用いる画像は実験4と同じもの(再現のきわめて難しい画像)とした。そして、この条件で作成する画像は、「実験3でかなり濃くスキャンした画像」(この時点では一番高い画質の画像)と比較することにした。

3-5-2 結果

実験の結果は次の通りである。

(1) 「組合せ番号10」および「組合せ番号19」でスキャンした画像(図9および図10を参照)の両方とも、実験3でかなり濃くスキャンした画像(図11を参照)よりも、読みやすくなっていった。具体的にはこうである。実験3でかなり濃くスキャンした画像では、各文字の線に濃い箇所と薄い箇所が混在しているために、離れて見ると、途切れた部分があるように見える。また、文字全体が幅の揃った線で構成されていないため(太線と細線が混在しているため)読みにくい。一方、前者2つでは、それらの読みにくくなる要素(「濃い線と薄い線の混在」と「太線と細線の混在」)が目立っては見られない。そして、全体的に文字が濃く見える。3つを並べると、一瞥して、文字の濃さの違いがはっきりとわかる。

(11) The U. S. Forces shall be authorized and other indirect firing over the reservoir, in order to impact areas. New impact areas will require

f. The Marine Corps does not guarantee minimum quantity of water from the Hansen reservoir

g. The Marine Corps shall retain the right of concurrence of the CINCPACFLT, in accordance with management, operation and maintenance of the base during duly declared national emergencies. The base is liable for payment of all rentals, claims, and other expenses subsequent to and during the period of such take-

図9：組合せ番号10でスキャニングした画像

(11) The U. S. Forces shall be authorized and other indirect firing over the reservoir, in order to impact areas. New impact areas will require

f. The Marine Corps does not guarantee minimum quantity of water from the Hansen reservoir

g. The Marine Corps shall retain the right of concurrence of the CINCPACFLT, in accordance with management, operation and maintenance of the base during duly declared national emergencies. The base is liable for payment of all rentals, claims, and other expenses subsequent to and during the period of such take-

図10：組合せ番号19でスキャニングした画像

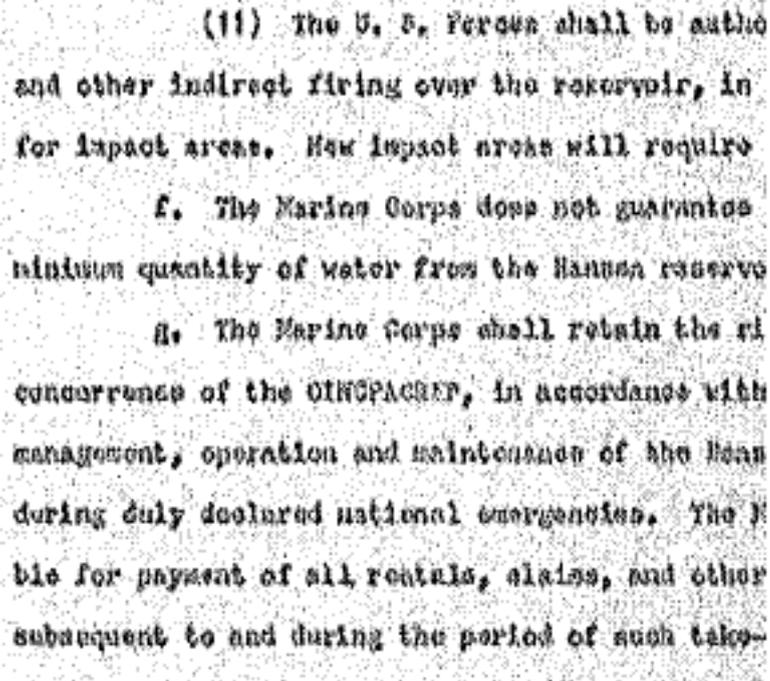


図 11：実験 3 でかなり濃くスキャンした画像

(2) 「組合せ番号 10」と「組合せ番号 19」の両プロファイルの出来を比較すると、1 点相違点が見られた。「組合せ番号 19」でスキャンした画像のほうが、文字とそれ以外の情報が明確に分けられていた。つまり、文字は濃くそして文字以外の部分（余白）は薄くなっていたのである。一方、「組合せ番号 10」でスキャンした画像は、実験 3 でかなり濃くスキャンした画像と比較すると、余白の部分の濃さについては同じであった。つまり、黒粒の多さは同じであった。

(3) 総合的にみると、「組合せ番号 19」のプロファイルが、文字と文字以外の情報を明確に区別するという点で、もっとも優れているということがわかった。同プロファイルを使えば、元画像の濃淡の影響を受けずに、文字の芯だけを正確にスキャンできるのである。これにより、1 つのマイクロフィルムをデジタル化するのに 1 回のスキャンだけですむようになると思われた。しかし、この時点では、同プロファイルの有効性は、1 つの画像に対して証明されたただけであった。「組合せ番号 19」が量産用として使えるプロファイルであるとはまだ断定できなかった。よって、多種多様な画像でテストする必要が出てきた。こうして、最後の実験をすることとなった。

3-6 実験 6 合格

3-6-1 目的と方法

実験 5 では、「組合せ番号 19」のプロファイルが、文字情報を明確にスキャンできるということが、証明された。たしかに、これはたった 1 つの画像に対するテストではあったが、同プロファイルの効力は高い水準で証明されたと言える。なぜなら、実験 5 で用いた元画像は、「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」という特徴を全て持っていて、なおかつ実験 3 で不合格となった、かなりスキャンの難しい画像だからである。このやっかいな画像でよい結果を出すことができたということは、他の様々な悪い画像でも上手く再現できると期待できるのである。あとは、実際に多くの画像で試してみ、実証するだけであった。実証されれば、量産用としても通用すると

結論付けることができる。

同プロファイルの有効性をさらに深く知るためには、かなり厳しいテストを行わねばならなかった。多種多様な読みにくい画像が含まれているリールで試さねばならない。そこで、筆者は、多くのマイクロフィルムを調べて、次のような特徴があるリールを選びだすことにした。1)濃い画像が多く含まれている。2)薄い画像が多く含まれている。3)「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」という要素を2つ以上もつ画像が、多く含まれている。とくに、最後の3)の特徴を持つ画像が数百以上入っているものを選んだ。なぜなら、3)のタイプの画像が不合格率を上げる最大の原因となっていたからである。

実験では、「組合せ番号 19」の能力を相対的にみるために、同プロファイルのほか、「実験 3 でかなり濃くスキャンしたプロファイル」と「実験 3 でかなり薄くスキャンしたプロファイル」についても、それぞれ1 リール分の PDF を作成した。

また、検査の精度を上げるために、あらかじめ、再現の難しいと思われる 120 のコマを厳選し、それらの番号を控えておいた。その 120 のコマは、「高い精度での再現は難しいが必ず合格にできる画像」「合格品を作るのが難しい画像」「合格品を作るのがほぼ不可能な画像」の3種類で構成されている。それらの画像のほとんどは、「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」という要素を2つ以上もつタイプである。つまり、不合格率を上げる最大の原因となっていたタイプである。スキャンの難しいこの 120 のコマで不合格が出なければ、全体（1867 コマ）でも不合格は出ないであろうと判断できる。また、この 120 のコマで高い再現性が認められれば、「組合せ番号 19」は、他のどのリールでも低不合格率を実現できると思えた。

3-6-2 結果

実験の結果は次の通りである。

(1)あらかじめ選んだ 120 コマのうち「濃淡両極端方式」では 12 コマが不合格となったのに対して、「組合せ番号 19」で作成した画像では、1 コマだけが不合格となった。ちなみに、その不合格の 1 コマとは、「日本語の新聞の切り抜き」(図 12 を参照)であった。これは、モニターに映し出されたマイクロフィルムの画像と比べると、明らかに画質が劣っていた。特に、画数の多い漢字では画質の劣化が顕著になることがわかった。「組合せ番号 19」でも限界があるということも証明されてしまったのである。しかし、逆に、この例外を除けば、「組合せ番号 19」で不合格を完全になくすことができるということが証明されたと言える。



図 12：日本語の新聞の切り抜き

(2) 全般的に、薄目の画像に対しては、「かなり濃くスキャニング」と「組合せ番号 19」が有効であった。ただし、a、o、d、p、e、s、m、g などのアルファベットの中白部分を潰さずに見えるようにするという点においては、「組合せ番号 19」のほうが優れていた（可読性が高くなっていた。図 13 および図 15 を参照）。「かなり濃くスキャニング」の画像では、これらの中白部分は、潰れてしまい、黒丸に見える場合が多かった（図 14 および図 16 を参照）。これにより、「組合せ番号 19」を使えば「かなり濃くスキャニング」の存在意義はなくなると結論付けられる。

1. A representative of the Naval Air
17 March 1959 that the following range will
Okino Daito (Also known as Rana Jia
From 1200 hours 18 March through 21
2. Request your office publish notice
media.
3. All departments which have been fu
are requested to advise their counterpart of
nation to all concerned.
4. A radio message will be dispatched
and ICAF for their information.

図 13：「組合せ番号 19」でスキャンした画像 1

1. A representative of the Naval Air
17 March 1959 that the following range will
Okino Daito (Also known as Rana Jia
From 1200 hours 18 March through 21
2. Request your office publish notice
media.
3. All departments which have been fu
are requested to advise their counterpart of
nation to all concerned.
4. A radio message will be dispatched
and ICAF for their information.

図 14：「かなり濃くスキャン」した画像 1

1. The 20 January problem involves a daylight maneuver area (for route and requested time 1). Once at the dam site a battalion anti place. At dusk the road march back to Cam

2. The ¹⁷ February problem involves a daylight training area. A daylight reconnaissance take place. After dusk a night problem in ONTOC alternate and supplementary position the conclusion of the problem a night road will take place.

REMARKS: The route and training area shown utilized by this battalion on 16 December training was considered very valuable. No Okinawa personnel or property areas. At approach closer than 300 yards to the dam.

図 15: 「組合せ番号 19」でスキャニングした画像 2

1. The 20 January problem involves a daylight maneuver area (for route and requested time 1). Once at the dam site a battalion anti place. At dusk the road march back to Cam

2. The ¹⁷ February problem involves a daylight training area. A daylight reconnaissance take place. After dusk a night problem in ONTOC alternate and supplementary position the conclusion of the problem a night road will take place.

REMARKS: The route and training area shown utilized by this battalion on 16 December training was considered very valuable. No Okinawa personnel or property areas. At approach closer than 300 yards to the dam.

図 16: 「かなり濃くスキャニング」した画像 2

(3)濃い目の画像に対しても、「かなり薄くスキャニング」より「組合せ番号 19」のプロファイルのほうが、うまくスキャンできていた(可読性が高く仕上がっていた)。これにより、「組合せ番号 19」

を使えば「かなり薄くスキャン」の存在意義はなくなると結論付けられる。

(4) 同 120 コマのうち 3 コマについては、「組合せ番号 19」で作った画像のほうが、元画像よりも読みやすくなっていた。元画像で見える「文字の滲んだ部分」が、「組合せ番号 19」ではなくなっていて、文字の芯だけが見えるようになっていた。

(5) 「組合せ番号 19」でスキャンした場合、再現の難しいコマだけを集めた 120 のコマでも不合格が 1 つしか出なかったので、全体（1 リール 1867 コマ）でも不合格はほとんど出ていないと想像できる。

3-7 収穫

6 回の実験をまとめると次のようになる。

(1) 不合格率を下げるには、文字と文字以外の情報を明確に区別して、文字の線の中心部分だけを正確にスキャンする（インクの滲みのせいで文字の線の幅が広がってしまっている場合、その滲んで広がった部分を無視して元の線だけをスキャンする）ことがもっとも大事である。今回、6 度の実験を経て、そのための最適なプロファイル「組合せ番号 19」を発見することができた。

(2) 「組合せ番号 19」の発見により、制作工数・時間を大幅に減らすことができた。実験 3 までは、2 回スキャンをして出来のよい方の画像を選ぶという方法で制作するしかなかったが、「組合せ番号 19」の発見により、1 回のスキャンだけで済むようになった。これは、質（画質）を上げることにより、量（制作工数・時間）を削減することができるようになったということの意味している。2000 コマ入っているマイクロフィルムを、30 分間でしかも高い再現性でスキャンできるようになったのは、画期的なことである。

(3) カーボン紙を使ってタイプ打ちしていた時代の資料は、全体的に、画質が悪い。マイクロフィルムに収められているそれらの資料を、画質を落とさず高速でスキャンするためには、RS200 などのマイクロスキャナーが必要となってくる。

(4) 「文字サイズが小さい」、「文字が薄い」、「文字が滲んでいる」、「文字以外の部分（余白）が濃い（黒粒が多い）」などの特徴をもつ画像については、RS200 のような高価な¹²マイクロスキャナーでも、プロファイルが適切でない場合は、スキャンがかなり難しくなる。

(5) 「組合せ番号 19」を発見できた理由のひとつは、同社が、高画質と低コストの実現を目指す実験を、RS200 の性能を知り尽くすための良い機会と前向きに捉えたことである。沖縄県公文書館と同社が、最後まで実験に対する探究心を失わず、建設的な議論を重ねることができたのも、互いに技術改革を求めていたからである。筆者は、現状に甘んじずイノベーションを目指すことの大切さを、今回の実験で、再確認できたと思っている。

(6) 成果を生んだもう 1 つの理由は、最初にハードルを高く設定したということである。562 ものマイクロフィルム（コマ数にして百万以上）を、「高画質で」「早く」「低価格で」デジタル化して公開しなければならないという厳しい制限があったからこそ、これまでになかった策を生み出していくことができた、と、筆者は考えている。

¹² 定価で 1,500 万円。

おわりに

筆者は、USCAR 法務局文書のデジタル化の業務を担当することになったとき、デジタル化するのであればユビキタスを実現しなければならないと考えた。つまり、いつでも、どこでも、手軽に資料が入手できるようにするということである。汎用性の高い PDF を選択したのも、その理由による。本文書の PDF は、数回に分けて、沖縄県公文書館のウェブサイト上で公開されていく。誰でも、特別な手続きなしで、PDF を閲覧できるようになる。端末機へのダウンロードも可能である。また、ネット上での学術情報提供の方法として、PDF ファイルがかなり普及している¹³ということも付け加えておきたい。

しかし、デジタルアーカイブズ資料をインターネット上で公開するという試みそのものは、すでに多くの公文書館や図書館などでなされているので、別段珍しいことではない。沖縄県公文書館のウェブサイトでも、すでに、多くの写真や映像が公開されている。では、今回の試みの独自の点は何なのか？ それは、次の3点にまとめられる。

(1) 通常のプロファイルでデジタル化した場合読めなくなるような画質の悪い画像を、読める水準を維持したままデジタル化する。つまり、「組合せ番号 19」のプロファイルを使うことによる不合格率の低減である。

(2) 1 リールにつき 2000 コマ近く収められているマイクロフィルム 562 本分を、4 ヶ月という短期間でデジタル化する（最終的に、同社は4 ヶ月で対応できると回答した）。早くデジタル化できるということは、工数や作業時間が少なくすむということであり、これはコストを抑えることにつながっている。単価/1 コマは 4 円弱となった。この価格がどれだけ低価であるかは、他と比較するとよくわかる。例えば、当館で実績のあるデジタル化の単価/1 コマは 40 円弱である。価格算出の条件が同じではない¹⁴ので、単純な比較をすることはできないが、USCAR 法務局文書のデジタル化が大幅な価格低減を実現したことは間違いない。また、関東イメージ情報業連合会が作成した 2007 年（平成 19）10 月時点での「電子ファイリング制作料金参考資料」によると、16 ミリメートルのマイクロフィルムを 400dpi の解像度でスキャンする価格は、1 コマ 60 円となっている¹⁵。

(3) USCAR 文書は「新しい」。「新しい」とは、USCAR が存在した時代の沖縄を知っている県民がまだ生存しているということ。

本業務を開始した頃は、専門業者に委託すればマイクロフィルムを画質の低下なしに PDF に変換することなどは簡単であると考えていた。しかし、現実はそうではなかった。最初に作成された PDF を見た時、画質の悪い画像の多さに驚き、様々な実験を重ねて画質を上げる必要があると思った。幸

¹³ 例えば、CiNii 収録の学術論文で全文検索の対象は「CiNii 本文 PDF」のみとなっている。また、「CiNii 本文 PDF」の利用頻度は高くなってきている。国立情報学研究所の大向一輝氏によると、CiNii では、Google の検索エンジンと連携するようになった 2007 年 4 月前後では、本文 PDF ファイルをダウンロードできるユーザ数に変化がないにもかかわらず、本文 PDF ファイルの利用回数が伸びたという。これは、ウェブ検索の過程で論文を発見し本文を入手するという状況がありえることを示している。参照文献：大向一輝「学術情報サービスのユーザモデルとファインダビリティ」『情報の科学と技術 58 巻 12 号』（国立情報学研究所 2008）pp.599-600。参照ウェブサイト：http://ci.nii.ac.jp/info/ja/articles/manual_keyword_full.html

¹⁴ USCAR 法務局文書のデジタル化の条件：16 ミリマイクロフィルムの PDF 化、1,056,026 コマ、モノクロ 2 値、600dpi。当館で実績のあるデジタル化の条件：16 ミリマイクロフィルムの PDF 化、120,000 コマ、グレースケール、300dpi。

¹⁵ 情報保存研究会『情報の保存と活用のための JHK ダイレクトリ 2011 年版』（2011 年）による。

い、筆者は、製造業で7年間不良品の改善や新製品の開発に携わった経験があるので、原因分析と対策に関するささやかなノウハウを持っている。しかも、納期とコストを絶えず意識しながら、計画の作成と実行、修正をすることができる。今回、その経験を、生かすことができた。「民間専門業者の技術力をいかにして最大限引き出すか」そして「制作期間をどれだけ短くできるか」、「画質を向上させつつ原価を低く抑えるためには、どれだけ簡素な工程にできるか¹⁶」がテーマであった。筆者は、問題点の整理と解決方法の探索には特に力を入れた。多種多様な「読みにくくなる現象」を分類・整理して、対策案を考えることがもっとも難しかった。しかし、それは、楽しくもあった。同社も、RS200の性能のアセスメントとより高度な活用という目的で意欲的に、またある意味楽しみながら実験に取り組んで下さった。同社のプロフェッショナリズムに敬意を表する。

また、2011年8月には、東京に出張し、東京都公文書館および国立国会図書館、アジア歴史資料センター、国立公文書館を訪問し、デジタルアーカイブズに関する多くの示唆深い助言をいただいた。ここに、謝意を述べさせていただく。さらに、筆者が作業を進めていく過程で、直属の上司として熱心に指導して下さい富永主幹と要所で有益な助言を与えて下さった宮城課長にも、お礼を申し上げたい。最後に、筆者が製造業勤務時代に厳しく指導下さったかつての先輩・上司の方々にも、お礼を言いたい。この時代に得た経験がなければ、今回の成果がでることはなかったと断言できるからである。

¹⁶ 画質が向上すれば工程を簡素化できるということがわかった。具体的には、スキャニングの性能を上げて画質を向上させることにより、スキャニング回数を2回から1回に減らすことができ、さらにその後の工程（目視選別）も削減できるようになった。