

画像データに関する調査報告書

平成11（1999）年3月23日

技術協力調整班

平成11年3月23日

電子図書館推進委員会

委員長 原 田 公 子 殿

技術協力調整班

班 長 田 屋 裕 之

「画像データに関する調査報告書」の提出について

標記について、技術協力調整班は、平成10年6月24日（水）以来、12回の会合をもち、電子図書館推進に伴う技術的課題について鋭意検討を重ねてきましたが、今般、別紙のとおり検討結果を取りまとめたので報告します。

検討経緯

○平成10年		第7回	10月29日（木）
第1回	6月24日（水）	第8回	11月25日（水）
第2回	7月9日（水）	第9回	12月25日（金）
第3回	7月22日（水）	○平成11年	
第4回	8月5日（水）	第10回	1月27日（水）
第5回	9月2日（水）	第11回	2月22日（月）
第6回	9月30日（水）	第12回	3月1日（月）

技術協力調整班

（管理職以外の職員名を含む一覧であるため、被覆）

画像データに関する調査報告書

電子図書館の実現のためには、電子図書館の「蔵書」の構築が必要である。電子図書館の「蔵書」は、二次情報と一次情報の大きく2つに分けられるが、特に一次情報については、平成10年5月に発表した「国立国会図書館電子図書館構想」第2章において、「一次情報の提供がインターネット等のネットワークを通じて可能となること」が電子図書館の長所として挙げられている。一次情報を電子化するに当たっては、さまざまな技術的な課題が存在する。特に既存の図書館資料の電子化については、問題点をクリアしつつ、電子化への対応が急がれる。本報告書は一次情報のうち、特に画像情報に限り、問題点の解消に向け技術上の検証を行った、その取り組みの結果を報告するものである。

印刷物など既存の資料の電子化を行う場合、画像原資料の形態や求められる作成データの質により対応方法や留意事項、使用する機器などが異なる。また、当館の資料は多種かつ大量であるため、それぞれの資料の特性に応じながらも効率的に電子化を行うことが求められる。そのためには、電子化の方法及び作成データの仕様について十分に検討し、当館の電子化を行う指針、もしくは基準を作成することが望ましい。

平成10年4月に電子図書館推進委員会の下に設置された技術協力調整班では、上記の目的のための調査・検討を行った。その対象としては、特に貴重書等の画像データ作成を進める図書部古典籍課からの要請を受け、画像データを対象に実施した。調査は、画像データ作成関連企業に対する技術条件についてのアンケート調査と、その回答に基づき、各社に電子化を依頼して作成されたサンプル画像データの検証からなる。

この報告書は、その調査と検討の結果を示すものであり、国立国会図書館としての画像データ作成基準作成への第一歩となるものである。

1. 当館における画像入力の実状と課題

1.1 画像データ作成対象

電子化を行う際、作成するデータをテキストデータと画像データのいずれにするかを判断する必要がある。両者の特徴を比較すると、以下の表1の通りになる。

表1 テキストデータと画像データの特徴比較

	対象となる蔵書(原資料)	入力方法	利点	欠点
テキストデータ	・文字	・手(パンチ)入力 ・OCR(光学式文字認識) ・出版データの活用	・文字列検索が可能 ・引用・翻訳等の再利用・加工が容易 ・データ量小	・原資料のレイアウト、イメージの再現が困難 ・原資料による技術的、コスト的な制限
画像データ	・絵または絵と文字の組み合わせ ・文字の形自体が調査研究の対象となるもの ・テキストデータの入力が困難なもの	・撮影→フィルムスキャナ ・ダイレクトスキャニング ・デジタルカメラ	・原資料のレイアウト、イメージを再現 ・ほとんどの原資料に対応	・入力方法、データの形式、利用方法等多くの未確定要因が存在 ・データ量大 ・検索・加工等困難

画像データの対象となる資料は、近代以前の活字資料のようにテキストデータの作成が難しいものや古典籍資料が挙げられる。特に古典籍資料は写本の筆遣いや古活字の形状を研究対象とする利用者が多く、また錦絵や本草、古地図といった絵を主体とする資料も含まれるため、電子化に際しては画像データとすることが極めて有効である。また、画像データはテキストデータと比べ、その作成に比較的成本がかからず、レイアウトを保存できることから、テキストの利点を必要としない利用を想定する資料については画像データとすることも考えられる。

1.2 当館における画像データ作成経緯

当館では、これまでも電子図書館の「蔵書」の構築のためにさまざまな画像データを作成している。しかし、それらは実験的な性格であったり、外部機関または業者が主体となって策定した仕様に基づいて、あるいは追従した形で原資料の所管部局レベルで作成したもので、その位置づけや仕様を館として総合的に検討したうえで作成したものではない。例えばモノクロ資料の場合、これまで一様に MMR 圧縮の TIFF データ(解像度 400dpi)を採用し、GIF データに変換して提供している。そのため、データ中の文字が判読しづらいもの、ファイルサイズやピクセルサイズが大きすぎるものといった、明らかにその利用に不都合をきたしているものもあり、見直しを行う必要があると考えられる(表 2 参照)。

表 2 当館における画像データの主な作成実績

年度	事業の種類	対象資料と数	画像形式と数
平成 6～7 年度	IPA 実証実験	貴重書全般約 130 点	PhotoCD 及び JPEG、約 7 千ファイル ※
同	同	明治期刊行図書約 25000 冊	MMR 圧縮 TIFF(解像度 400dpi)、約 325 万ファイル ※
同	同	昭和期劣化資料約 2500 冊	MMR 圧縮 TIFF(解像度 400dpi)、約 38 万ファイル ※
同	同	国内刊行雑誌 24 誌	MMR 圧縮 TIFF(解像度 400dpi)、約 84 万ファイル ※
同	同	国会審議用調査資料 264 冊	MMR 圧縮 TIFF(解像度 400dpi)、約 5000 ファイル ※及びテキスト
同	同	憲政資料 3433 点	MMR 圧縮 TIFF(解像度 400dpi)、約 3 万 3 千ファイル ※及びテキスト
平成 7 年度	アジア文献情報	アジア資料(図書 1361 冊、雑誌 18 誌、新聞 7 紙)	
同 7～8 年度	児童書基盤系システム	児童書約 7000 冊	カラー…PhotoCD 及び JPEG、約 1 万 9 千ファイル JPEG512×718 ピクセル モノクロ…MMR 圧縮 TIFF(解像度 400dpi)、約 90 万ファイル
平成 9 年度	展示会(BBCC 協力事業)	伊能図 43 枚	非圧縮 TIFF、約 80 ファイル
同	古典籍資料電子化	古典籍本草等 12 点	PhotoCD、約 3200 ファイル

※表示の際にサーバー側で GIF データに変換している。

また、これまでの画像データの作成記録や仕様書等は、集中管理されておらず、情報の共有化もなされていないため、経験や問題点の共有がなされないことが懸念され

る。そのため、今後電子図書館の「蔵書」として利用され得る画像データを作成していくためには、標準的な入力仕様、指針もしくは基準の作成とその全館的な共有が必要となる。

その一方、最初に画像データの作成を始めた平成 6 年度当時と比較して、画像データの表示が容易になっている。画像データの表示に用いる PC やネットワーク環境も著しい向上を見せている。新たな規格の画像データ形式についても考慮が必要である。

2. アンケート調査の実施及び結果

2.1 概要

画像データ作成基準の作成に向け、図書館資料の電子化に適する画像データフォーマットやその設定について把握するため、まず最初に画像データ作成関連企業に対するアンケート調査を行った。このアンケートでは、画像データ作成時に必要な技術条件（入力条件、画像フォーマット、圧縮方法、出力条件等）について、以下の 3 点を中心に尋ねた。

- ・各種資料からの電子化への対応
- ・入力仕様
- ・電子化に当たっての判断基準

2.1.1 調査対象資料

今回の調査では、その範囲を当館の代表的な資料形態であるフィルムと印刷資料に絞って行った。

2.1.2 協力企業

本調査の趣旨に対し賛同を得た画像データ作成関連企業 17 社の協力により実施した。企業名は以下のとおりである。（会社名の五十音順。）

- アシストマイクロ株式会社
- 株式会社アルファインターネット
- NTT株式会社
- 株式会社紀伊國屋書店
- 株式会社経葉社
- コダック株式会社
- 株式会社ジェイ・アイ・エム
- 株式会社写真化学
- 新日本製鉄株式会社
- 大日本印刷株式会社
- 大日本スクリーン製造株式会社
- 凸版印刷株式会社
- 株式会社ニチマイ
- 日本写真印刷株式会社

- パイオニア株式会社
- 富士写真フイルム株式会社
- 丸善株式会社

2.1.3 調査方法

質問票によるアンケート方式にて2回実施した。1回目は、画像データの電子化に関する基本的な事項について、2回目は1回目のアンケート結果を踏まえ、より専門的な事項について追加質問を行った。

2.1.4 調査期間

平成10年8月26日～平成10年9月25日

2.1.5 調査結果の集約

回答の取りまとめに際しては、データの公平性を保ち、また企業による先入観を取り除くため、企業が特定あるいは推察される情報を割愛した上で、各項目毎に回答の記述順序を変更した。また、アンケートへの回答をそのまま取り入れるのではなく、代表的な回答や今後の当館の電子化にとって参考となる回答について編集、分析を行った。

2.2 調査結果

2.2.1 フィルムからの電子化への対応

(1) 企業別対応可能範囲

企業によって対応可能なフィルムに多少の違いはあるが、モノクロフィルムでは98コママイクロフィッシュ、16mmマイクロフィルム、35mmマイクロフィルム、カラーフィルムでは35mm、4x5判、8x10判が、どの企業もほぼ対応可能との回答であった。

(2) 電子化に伴う取り組み方

①モノクロフィルム

フィルムの違いだけでなく、フィルムに撮影された資料や利用の形態に対応する電子化方法を取る必要がある。撮影された資料が活字中心の場合は二値変換、古典籍資料など筆跡や墨の濃淡等の情報も含めて表現する必要がある場合にはグレースケール変換もしくはカラー画像変換が有効である。但しグレースケール変換には画像サイズが大きくなる傾向があり、その点の考慮が必要である。

②カラーフィルム

カラーの場合は原資料の色を可能な限り忠実に再現することが求められる。色は出力機器やアプリケーションに左右されることが多いため、標準化されたカラーマネジメントの仕組みを利用し、電子化を行うことが必要である。ICC規格で標準形式として定義されているICCプロファイル（カラー空間を記述するファイル）を使用す

ると、異なるプラットフォームや出力機器、ICC 準拠のアプリケーションにおいてカラーを正確に再現しやすく、カラーマネジメントが容易になる。

(3)使用する電子化機器

①モノクロフィルム

フィルムスキャナを使用する。

②カラーフィルム

高解像度データにはドラムスキャナ、高精細フラットベッドスキャナを使用、通常の画像データについてはフラットベッドスキャナを使用することが多い。

(4)フィルムからの電子化に当たって特に留意すべき事項

①入力準備に関すること

- ・ 利用形態が定まらない内に電子化を進めると、品質面での問題や無駄なデータ変換作業等が必要になるため、利用形態を考慮した電子化を行う。
- ・ 原資料がフィルム上に正しく撮影されているか、データと余白の割合、傾き、発色等について確認する。
- ・ デュープフィルムを必ず作成し、オリジナルフィルムを保全する。

②データの入力に関すること

- ・ 原資料を忠実に再現するため、実際に原資料を確認した者による明確な入力指示が必要である。
- ・ 利用面、保存面をともに考慮した可能な限りの高画質を実現する。
- ・ できるだけ標準的なフォーマットを採用する。
- ・ フィルムによる解像度の違いを確認する（例：ネガよりリバーサルフィルムの方が画像データの再現性が高い）。

③データの補正に関すること

- ・ フィルムに付着したキズ、ノイズを除去する。
- ・ カラーフィルムの経年的な劣化や、破損、汚損等を復元する。

なお、可能であれば以下の項目に対処する。

- ・ フィルムの発色の違いを補正する。
- ・ カラーマネジメントを行う。
- ・ 色調、階調性、解像性、シャープネス、粒状性、安定性を統一する。

④作成データの保管に関すること

- ・ 今後想定される用途に対応するため、標準的なフォーマットを使用するとともに、十分な大きさの画像データとして保管する。

⑤その他

- ・ 高精細なデジタルカメラへの対応を考慮する。(近年急速に高精細画像に対応する製品が出されてきている。現時点では解像度の点でフラットベツトスキャナ、ドラムスキャナには及ばないとされる。)
- ・ 今後増加するであろうフィルムの新規格 (APS 等) への対応を考慮する。
- ・ 作業工程の簡略化、作業時間の短縮を図る。

(5) フィルムの画像汚れを自動的に補正する技術

スキャン時に発生するモアレや裏面の写り込みなどの補正作業は、スキャナのノイズ除去機能や画像処理ソフトを使用することで、ある程度自動化できる場合がある。

ただし、原資料や入力データにより対応が異なるため、一括して自動補正を行うと必要以上に補正される恐れがあり危険である。機能的に自動補正が可能であっても、実行して問題がないか事前に見極め、可能であると判断した作業以外は手作業で行うべきである。

(6) フィルムスキャン時にフィルムが破損する可能性の有無

スキャン時に使用する機器やその設定によってはフィルムが破損する可能性がある。フィルムが直接接触する箇所はクリーニングを行う。また直接接触することを避けるためにフィルムと機器の間に透明なシートを補うなど間接的にスキャニングを行うことも有効である。

(7) デュープフィルムからのスキャニングにおける画像劣化の度合い

デュープフィルムはオリジナルフィルムを保全するために作成するのだが、解像度、画質、コントラスト、発色等において若干の劣化が発生する。デュープ作成時には最小限の劣化に止めるよう注意し、万一目視できる程度の劣化が生じた場合は補正することも考慮する。

2.2.2 印刷資料からの電子化への対応

(1) 印刷形状への対応

①一枚もの

直接スキャニングができる場合はフラットベツトスキャナでスキャニングする。
できない場合は、フィルムに撮影後、そのフィルムをスキャニングする。

②巻物・大型資料

大きさに応じたコマ数でフィルム撮影後、そのフィルムをスキャニングし、必要に応じ画像接合処理を行う。

③裁断可能な製本資料

基本的に一枚ものと同様。定型で一定強度ならば業務用高速ドキュメントスキャ

ナ、フィーダー付きスキャナを使用し、効率的なスキャンニングも可能である。

④裁断不可能な製本資料

フェイスアップスキャナによるスキャンニングまたはフィルムに撮影後、そのフィルムをスキャンニングする。

(2) 中間媒体の使用の有無、使用する場合その想定媒体

原資料の劣化が顕著であったり、見開きにするなどのストレスをかけることができない場合は、原資料を傷めないためにフィルム等の中間媒体を使用することが望ましい。また、資料が大型でかつ高精細、超高精細なデータが必要な場合等、現状の電子化機器で対応できない場合においても、今後の新たな技術レベルでの再スキャンニングを行う際に再撮影を不要とするよう中間媒体を作成した方がよい。いずれの場合も撮影時に直接多量の光量を照射することが可能かどうかの判断が必要である。中間媒体としては、フィルムを使用することも可能であるが、電子化を前提に使用すると、画像が劣化すること、電子化の工程が特別なものとなることを考慮に入れる。

(3) 印刷資料からの電子化に当たって特に留意すべき事項

- ・ 目的に応じたデータの作成や取り込み時の資料破損・汚損の防止のため、資料の形状、求める画像品質、スキャナー効率、使用する機器や治具等について事前に綿密な打ち合わせを行っておくことが望ましい。
- ・ 絵柄や平網等の網がけされている部分がある場合、入力時モアレが発生することがある。
- ・ 白黒二値で取り込む資料で、レイアウトや装丁に関する情報が必要ないデータの場合は、テキスト化についても配慮する。テキスト化によりデータサイズの縮小、内容の検索などの利便性の向上が図れる。
- ・ カラー画像を大量に作成する場合は、同資料内にばらつきが発生しないようカラーマネジメントを行う。
- ・ 必要があれば画像補正を行う。
- ・ フィルムへの複写時、左右の照明強度、画像の歪み、撮影機材特性を考慮する。特に製本資料の電子化の場合、電子化機器へのセッティング状態の違いにより中央部分と端でフォーカスや階調再現性が異なる。
- ・ 保存や流用、再編集を考慮した十分に高い解像度を保持する。
- ・ 利用に十分供するよう実物に近い色数等を考慮する。

(4) 破損資料または破損の恐れがある資料を電子化する具体的な保護措置

- ・ 原資料に接触しないよう、一旦フィルムを介してスキャンニングしたり、ガラス、アクリル板、クリアケースなどを用いる。
- ・ 入力環境を恒温恒湿にする。
- ・ 照明は熱の出ないものにする。

- ・ 資料の損傷によってはL字形の撮影架台を使用する。
- ・ 可能な場合は事前にしわ伸ばし・補正用テープでの補強等を施す。

2.3 入力仕様

電子化を行う際、その対象となる原資料の状態や利用形態に応じて適切な画像ファイルのフォーマット、各入力条件の値、格納媒体、出力条件などが異なる。そのため、入力仕様にこれらを要件として明示しておく必要がある。ここでは電子化の仕様を作成する際に必要となる要件について示す。

2.3.1 画像フォーマット

(1) TIFF 形式における各種オプションの画質や運用面の相違

①LZW 圧縮、その他圧縮と非圧縮

・ LZW 圧縮

画像の劣化がないが、圧縮率が低いため、運用面において実用的とはいえない。研究用、保存用には有効である。また、対応していないソフトウェアがあるので事前に確認する。

・ 非圧縮

カラー画像の場合、マスタデータとして最適である。二値画像の場合は同じ可逆圧縮である MMR 圧縮が利用できる。グレースケール画像については、運用面においては JPEG による圧縮処理が有効である。

②インテル配列とモトローラ配列

画質面では問題ない。モトローラ配列は Macintosh でのフォーマット形式、インテル配列は DOS、Windows でのフォーマット形式である。通常は、インテル配列による納品が大半だが、画像処理に使用する端末に Macintosh が多い関係上、モトローラ配列も多い。

③マルチページとシングルページ

一つの TIFF ファイルに複数の画像を格納するかどうかの設定である。画質上は問題ない。ただし、100%対応が出来ているビューワーは少ない。シングルページはシーケンシャルなデータ・資料や運用面で利便性がある。容量が少ないためネットワークで使用しやすいが、ファイルの数が多くなるため管理しづらい。通常はシングルページである。マルチページにした場合、ページめくりができるが、データ量が大きくなるので運用上問題が発生する可能性がある。データの整理、リンク情報の付与など情報単位が細分化され階層化されている資料に関しては利便性がある。

④その他一般的に用いられるオプション

TIFF 形式で使用されるタグに非標準なものが使用されている場合、TIFF 画像を表示するソフトによりエラーとなる可能性がある。

(2) フォーマット変換の問題点

① JPEG の圧縮

アプリケーションが行う圧縮の範囲を目安として段階付けたもので、単位という概念ではない。アプリケーションによって 10 段階から 100 段階ある。1/20 以上圧縮すると劣化が肉眼で判別できるので、それ以下の圧縮率が望ましい。

② 保存用から提供用といった画像フォーマットの変換において、画像劣化などの問題が生じる組み合わせ

a) JPEG への変換

形式変換による劣化でなく、圧縮による劣化がある。非可逆圧縮で劣化は蓄積するため、2 度以上の圧縮は画像が大幅に劣化する。提供用画像形式として JPEG 形式を選択する場合、高圧縮率の場合には画像中のエッジ部分におけるノイズの発生は避けられないが、低圧縮率を選択した場合にはノイズの発生を極力抑えることが可能である。

・ TIFF→JPEG

オリジナルを TIFF、提供用を JPEG とする組み合わせはよく用いられている。JPEG の 1/10 程度の圧縮であればほとんどの劣化は視覚認識できない。

・ PhotoCD→JPEG

画像の輪郭が劣化したり、文字の部分に関しては画像にシャープ処理が必要なものが発生する。また、複数の解像度のデータを持つため、どれを用いるか注意する。

b) PDF への変換

階調のあるデータからの場合は文字や画像が滲むなどの劣化が生じる。

PDF のイメージデータの大きさは縦横共に 3200 ピクセルまでしか許されていないため、元 TIFF データがそれを越える場合は、イメージデータを縮小して縦横のピクセル値を 3200 ピクセル以下にする必要があるため、縮小時に画像の劣化が起こる。PDF では画像を圧縮するために LZW、JPEG、CCITT グループ 4 (TIFF-G4 圧縮)、ZIP、Run Length 他複数の圧縮技術を複合させているため画像の劣化は避けられない。

・ TIFF→PDF

二値画像の場合、MMR 圧縮ができるため、画質の劣化は起こらない。

・ PhotoCD→PDF

LZW 圧縮、JPEG 圧縮のいずれかが使用可能である。LZW 圧縮を使用すれば画質の劣化は起こらないが圧縮効果が低い。JPEG 圧縮を使用すれば高圧縮の場合、画像の劣化が起こる。

・ FlashPix→PDF

LZW 圧縮、JPEG 圧縮のいずれかが使用可能である。LZW 圧縮を使用すれば画質の劣化は起こらないが、JPEG 圧縮を使用すると画像の劣化が起こる。また、ダウンサンプル (解像度の設定) によっては劣化がある。

2.3.2 入力条件

画質の面からは、高解像度、多ビット、非圧縮が望ましく、また、データ量削減の観点からは、低解像度、低ビット、高圧縮率が求められるため、どの辺りで折り合いを付けるかが問題である。そのためには最終的にどのような利用目的、利用対象とするのかを明確に設定し、それに応じて各値を決定することが必要である。以下の値はあくまでも目安であり、実際には原資料の特質、利用形態等を含めて判断する。

(1)解像度

①保存・研究用

カラーの場合、原資料に対し 400dpi 以上、4000x5000 ピクセル程度（4x5 カラーポジフィルムを想定した場合）が適当と思われる。白黒二値の場合は、300～400dpi 程度が適当であると思われる。

②プリントアウト用

一般的なサービス用としては 300～400dpi 程度である。ただし、グラビア印刷等高品位画像であれば 1200dpi 以上という回答もあり、利用目的に応じて必要な値を設定する必要がある。

③ネットワーク提供用

72dpi、1000x800～1500x1000 ピクセル程度。実際には画面の表示サイズ等も考慮した上での設定が必要である。また、利用者のネットワーク環境はさまざまであるため、解像度の高さよりもサイズを小さくすることが求められる。ネットワークの伝送容量が低い環境にある利用者にストレスを感じさせない程度とすることに心掛けるべきである。

(2)階調、発色数

保存・研究用、プリントアウト用、ネットワーク提供用それぞれにおいて、階調はカラーの場合 RGB 各色 8 bit256 階調、1670 万色が一般的である。モノクロの場合、二値を用いるか、グレースケールを用いるかは対象資料による。

(3)圧縮方法、圧縮率

①保存・研究用

非圧縮が望ましい。

②ネットワーク提供用

カラーは JPEG 圧縮で約 1/10～1/20 程度、高品質を望む場合は、1/10 前後がよく用いられる。白黒二値は MMR 圧縮で約 1/10 程度が適当との回答が多い。

2.3.3 格納媒体

(1) 格納媒体

用途が保存用か提供用かによって大別することができる。

①磁気テープ (DAT を含む)

低コストで高速な記録が可能である。保存より運用に有用で大容量データの扱いに優れる。構内 LAN、Web サーバでの利用が想定される。しかし、巻き戻しや頭出し処理に時間がかかるため、任意の位置のデータを読み出す処理が遅い。データの検索、閲覧に時間がかかり、リアルタイム検索を行う電子図書館の記憶媒体としては不向きである。

②CD-R

世界的に使用され、パソコンにも標準でドライブが装備されているため、汎用性が広い。一度だけ書き込み可のためデータ改変の心配がなくデータ保持で優れる。耐用面でもメーカー保証 100 年のものもあり、実用範囲である。スタンドアロン、CD サーバでの利用、保存媒体としても有効である。価格も比較的安く、大容量の格納を実現できる。また、読み取りはレーザーにより非接触で行われるため、読み取りによるメディアの劣化が起こりにくい。しかし、情報を記録する金属蒸着膜が化学反応して劣化するとも言われている。

③DVD-R、PC-RW

画像データ保存での CD-R の代替メディアとして発展する可能性があると考えられる。PC-RW はソニー、フィリップス、ヒューレット・パッカードによる書き換え可能 DVD 規格。両者に互換性のないことが問題とされる。

④MO

一時的な保存、搬送用として利用。格納媒体としては書き込み可能なところに問題がある。

⑤磁気ディスク

任意の位置のデータを瞬時に読み出せ、電子図書館資料の検索、閲覧が瞬時に可能である。しかし、高価なハードディスクを大量に用いるため、他の媒体に比べコストがかかる。別の媒体で作成したデータを、提供用に保持する目的で使用され、保存用としては不適であるが、データ提供用には不可欠である。

(2) 格納形態

①ボリューム名

目的による任意の設定でよい。ただし DB による別管理が必要となる。

②ファイル名

ISO9660 Level 1「英数字 8 桁+.3 文字の拡張子」形式を推奨する回答が多い。

③ディレクトリ構造規則

シンプルで意味のある構造とし、無意味に深くしすぎないようにする。

(3)受注の割合から見た各種拡張 CD-ROM 規格の普及率

ISO9660 Level 1、ISO9660 Level 2 が一般的である。

(4)媒体ごとにデータ破損、劣化が想定されるケースとその対処方法。または図書館でその媒体を保存する際の注意点

破損・汚損、劣化を防ぐためには、物理的な要因に対しては温度や湿度、埃などを考慮した空調設備のある専用保管ルームでの保管が望ましい。媒体の性質による劣化の発生や閲覧時の不用意または故意によるデータの破損、改変等からデータ内容を保護するためには必ずバックアップを複数（出来れば媒体を変えて）作成したり、データにプロテクトをかける。また、コンピュータウイルスへの対応としてセキュリティ措置を取る必要がある。

2.3.4 出力上の想定

電子化を行う場合、その利用目的や利用環境を想定する必要がある。その上でそれぞれの場合どのような提供方式を採るのか、最適な機器やその設定はどうあるべきかについて検討する必要がある。

(1)推奨する表示方法(ピクセル、発色数など)

sRGB 規格により調整したモニタに表示する。sRGB 規格に調整されたモニタに表示したとき、フィルムと同色の色設定が行われる。また、著作権保護等の目的から、電子透かし等を入れることも考慮されるべきである。

(2)利用環境

①スタンドアロン

大きなサイズのデータ表示が可能であるため、高精細データ表示のための環境として適切である。高精細データを有効に表示するため、ディスプレイは高精細データに対応し、できるだけ画面サイズの大きいものを考慮すべきである。

②構内 LAN

インターネット等で用いられる通常の公衆回線と比較して、通常高速の伝達が可能であるため、ある程度大きなサイズのデータでも対応できる。

③インターネット利用

回線速度が遅い場合が多く、また、さまざまな環境の利用者が想定されるため、

ピクセル数や発色数を小さくするなど伝送効率を重視した表示方法とする。

(3) 表示に想定すべきソフトウェア

構内 LAN やスタンドアロンにも WWW ブラウザを表示ツールとして利用すると、構内と構外を意識することなく利用可能となるという利点がある。また、コンテンツのコピーやプリントアウトの防止のため、個人ユーザーの認証用として電子図書館用ブラウザの別途開発の検討も考えられる。以下の閲覧ソフトは、アンケートで代表的なものとして示されたものである。

- ・ Netscape Navigator
- ・ Internet Explorer
- ・ AcrobatReader (PDF を想定)
- ・ Photoshop
- ・ Photoshop Pro
- ・ LivePicture (FlashPix を想定。但し LivePicture Image Server があれば不要)
- ・ Picture It (FlashPix 想定)
- ・ Imaging
- ・ Tau Scope

(4) プリントアウト手法

プリントアウト手法で主なものとしては、銀塩方式、昇華型転写方式、レーザープリンタ方式がある。画質を優先する場合は銀塩方式、ある程度の画質で速度を優先する場合は昇華型転写方式、速度及びランニングコストを優先する場合はレーザープリンタ方式を採用する。よって、写真と同程度の高画質を求める場合は銀塩方式、高画質が要求される貴重書や絵図などは昇華型転写方式、文字データがほとんどのものやそれほどの高画質が要求されない画像データはレーザープリンタが適当である。

(5) プリントアウトに求められる要件

1. データのレベル (保存用データか、提供用データか)
2. プリントサイズ (定型用紙内に収まるか)
3. プリント管理

この他に留意することとして、不正なプリントアウトを防止するために電子透かし等を入れる技術の利用、カラーマネジメントの実施、対応するソフトウェアの選択などが挙げられる。

(6) 電子化されたデータの品質判断の方法、検証方法

原資料を忠実に再現しているかが重要である。確認方法としてはディスプレイ表示、プリントアウトの区別に関わらず、基本的に目視による確認が一般的である。ただし、目視は検証者のレベルによってばらつきが生じる可能性がある。そのため、標準化団体によるテストチャートを使用するのもよい。テストチャートには画像電子学会「デ

デジタルスチルカメラ用テストチャート」、ISO-TC42-WG18「Chart for electronic still picture camera resolution measurement」の2種類が挙げられる。

2.4 電子化に当たっての判断基準

ここでは当館側があらかじめ電子化の想定条件を示し、それに対する企業の判断を求めた。想定条件は、下記のとおりである。

<想定条件>

a) 入力対象資料

ルビのあるモノクロ活字資料(A5判程度)及び文字の細かい古文書(B5判、筆書き文字10ポイント程度)

b) 提供方法

・研究者向きの高精細表示及び電子データの保存

モノクロデータ：原資料に対し400dpi相当にて、TIFF非圧縮で電子化。

カラーデータ：原資料に対し4000x5000ピクセル相当にて、TIFF非圧縮で電子化。

・構内LANにおける一般利用者利用、WEB上での一般利用者への提供

モノクロデータ：TIFF画像に対し、MMRもしくはMH圧縮を行い、パソコンの全画面に最大限表示可能なピクセル数(例えば、1024x768)に変換して提供する。また、グレースケールの階調を変えて提供する。

カラーデータ：TIFF画像をJPEG画像に変換し、パソコンの全画面に最大限表示可能なピクセル数(例えば、1024x768)で、1/10以下の圧縮率で圧縮する。必要に応じ圧縮率を変えて提供する。

・サムネイルの表示。

・利用者へのプリントアウトサービスの実施。(研究者向け高精細画像を用いる)

2.4.1 Web上での画像表示

(1) 文字を読むのにふさわしい解像度か

文字が中心となる資料においては、利用者がそれを読むことができるかどうかで最低限の判断基準となる。想定した解像度については、ほとんどの企業がモノクロデータ、カラーデータ共に問題ないとの意見であった。ただし、ルビを含めて考えた場合は、600dpi以上の解像度が必要であるとの意見もあった。また、モノクロデータの場合、データ量が大きくなるとの指摘があった。

(2) ネットワーク負荷が適切なデータ量か

ネットワークによる提供の場合、ネットワーク回線の伝送容量や構成等によって伝送・表示速度が異なる。そのためより多くの利用環境に対応できるデータ量に押さえ、利用者にストレスを感じさせないことが大切である。想定した圧縮方法、圧縮率にて

圧縮を行った場合については、適当であるとの意見が大半であった。

(3) サムネイル表示の適切なサイズ

サムネイルについては、絵や図、全体イメージと、文字のみで構成される文書データでは、サイズを変える意見が多かった。提案されたサイズにおいては、前者が 90x90～200x300 ピクセル、後者が 100x100～100x200 ピクセルの範囲とした回答が主であった。全体としては、絵や図、全体イメージより、文字のみの方のデータサイズを大きくする傾向が見られた。

2.4.2 構内での利用

(1) 来館利用者向けのデータと保存用データとを同一のものとする点

この点については意見が分かれた。反対意見としては、保存用データはサイズが大きくなるので利用に適さないこと、来館者のニーズに対応できるよう何種類か圧縮率を変えたデータを用意する方がよいことが理由に挙げられた。賛成した企業が挙げた理由としては、保存データと同様の高精細データが提供できること、変換作業が伴わず経済的で、作成から公開までの時間が短縮できることであった。これと同時にバックアップやセキュリティ設定、電子透かし等の不正防止措置等を行うことも指摘があった。

(2) プリントアウトサービスに用いるデータを研究者向け高精細画像とする点

これについても意見が分かれた。反対意見としては、プリント速度やプリンタの精度や、高精細データ以外の需要の可能性を考慮すると、必ずしも研究者向け高精細画像とする必要はないというものであった。

(3) 館内 LAN における一般利用者と Web 上での一般利用者は、基本的に同一のものとし、階調あるいは圧縮率を変えることに止める点

これも賛否が分かれた。反対意見は(1)と同様、予め何種類か階調、圧縮率を変えたデータを準備しておくという別の方法を挙げるところが多かった。また、Web 上での一般利用者向けには、館内 LAN における利用者向けより高解像度のデータとしたり、多少画像処理をして見栄えを良くするとの意見もあった。また、電子透かし等の不正利用防止措置を取ることを考慮すべきとの指摘もあった。

3. サンプル画像データの検証

3.1 概要

アンケート調査によって、多くの候補のなかから図書館資料の電子化に適すると思われる画像データフォーマットやその設定について、大枠については把握することができた。しかし、より最適なフォーマットとその設定を見出すためには、さらに具体的な調査が必要となった。そこで、実際にサンプル画像データを作成し、その比較・検証を行うこととした。

サンプル画像データの作成にあたっては、アンケート結果を踏まえ、電子化に適すると思われる画像データフォーマットを複数選定し、それぞれのフォーマットで作成条件をさまざまに設定した上、協力企業に依頼した。

調査は当館で所蔵するフィルム媒体からの電子化に限定して行った。当館で電子化を最優先する資料が、その形態や資料保存上、まず撮影によってフィルムを作成し、そのフィルムをスキャナで入力することが前提となるものが多いことがその理由である。

今回の検証は、フィルムに限ったが、この検証によって最終的な生成物である画像データのフォーマットを定義することで、原資料(紙媒体)からの電子化の基準も見出すことが可能になると思われる。

3.1.1 フィルム(資料)の選定

当館で所蔵するフィルムの形態は、ネガとポジ、カラーとモノクロ、リールとフィッシュまたはスライド、16mm と 35mm、ブローニー版と 4x5 または 8x10 など多種多様である。フィルムの選定に際しては、閲覧による摩耗でその特性が画像データから見えにくくなる恐れのあるポジのロールフィルムおよびマイクロフィッシュ、数量が少ないカラーネガフィルムを除外し、以下に挙げる特徴的な資料(被写体)と組み合わせ、10 通りのグループを作った。

(1) 幕末から明治初期の雑誌・新聞

古典籍課所管ですべて準貴重書に指定済みである。閲覧にはマイクロのみで対応しているが、利用頻度は高い。単色刷りで木版刷りと活版印刷がある。原資料の大きさは最大で B4 判程度で、現在の新聞よりもコンパクトなサイズである。フィルムは 16mm と 35mm のネガフィルムがある。

検証用としては原資料が木版刷りの 16mm と 35mm のネガフィルムをピックアップした。各社 1 リールずつとしたが、足りなかったため同様の資料形態を持つ木版刷りと古書のフィルムを補った。以下、16mm のものを b 群、35mm のものを w 群と呼称することにする。

(2) 本草資料

当館有数のコレクションとしてマイクロ、原資料ともに利用が多い。彩色写本が多いが、無彩色写本、木版刷りもある。

検証には無彩色写本の 35mm ロールフィルム(以下 h 群)、彩色写本のブローニーフィルム(以下 k 群)、同じく彩色写本の 4x5 フィルム(以下 s 群)を候補とした。

(3) 明治期刊行図書

約 30 万冊のすべてが 16mm のモノクロフィルムでマイクロ撮影済みである。木版刷り、活版刷り、単色、多色と雑多であるが、検証には他の検証資料との類似を避け、活版刷り資料のもの(以下 m 群)で統一した。

(4)昭和前期刊行図書

他の時代の資料と比較して紙質が悪く、劣化や破損が著しいため、図書閲覧課所管資料の中では優先的かつ計画的に 16mm フィルムによるマイクロ撮影を推進している。すべて活版印刷であるが、電子化の要望が高いことと、m 群との比較のために検証候補とした。なお、同フィルムには撮影年代と撮影業者が異なる 2 種類が存在するため、比較できるように別グループとした。一方を a 群、もう一方を r 群とする。

(5)錦絵、(6)古地図

どちらも非常に閲覧者の利用が多く、かつ利用による褪色や破損の可能性がきわめて高い資料である。早急な電子化が望まれるため、適切な画像フォーマットなどの策定も急ぐ必要がある。検証用としては、錦絵は 4x5 カラーポジフィルム (n 群)、古地図は 8x10 カラーポジフィルム (以下 c 群) とした。

3.1.2 サンプル画像フォーマットの選定

画像フォーマットの選定は、アンケートの結果を参考にした。

(1)モノクロ画像

①TIFF

モノクロ画像データのフォーマットは、二値入力 of TIFF で MMR 圧縮したものがよい、とする意見は非常に多かったが、入力時の解像度設定には諸説があり、複数の設定で検証する必要があった。また、過去に作成された二値入力の TIFF データの中には文字がかすれて判読できないものがあり、多値にも検討の余地があった。入力解像度の指定は、アンケートの回答をもとに 200、300、400(二値は 600 も)dpi とした。圧縮の有無については、圧縮による画像劣化が発生しないとの回答に基づき、特に圧縮なしの指定はしなかった。ほかに、二値は Revision6.0 Class F、インテル配列、シングルページ、サムネイル保存なしの設定で、多値は Revision6.0、インテル配列、シングルページ、LZW 圧縮、サムネイル保存なしの設定で統一を図った。

②GIF

アンケートの回答には見られなかったが、圧縮率が高く、WEB ブラウザで表示可能ということもあり、選択肢に加えた。扱える色の組み合わせが 256 色までというのも、多値の一般的な入力階調である 256 と一致し、選択肢に加えた一因でもあった。

(後にこの考えは誤りであることが判明した)。諸設定は、GIF89a 形式、ノンインターレース、サムネイル保存なしとした。作成方法としては、一般的な作成手順に従い、TIFF データからの変換による作成を指定した。

③JPEG

不可逆圧縮という欠点を差し引いても、高い圧縮率とインターネットなどにおけ

る普及率の高さは魅力的で、利用データの筆頭候補に挙げられる。問題となるのは圧縮率の程度であり、活用するのであれば、圧縮率について何らかの基準や目安が必要である。これらについてはアンケートに決定的な意見がなく、議論を重ねた末に、代表的な画像処理ソフトで今回の検証にも用いた Photoshop の圧縮設定に倣った「高画質(低圧縮)」「標準画質」「低画質(高圧縮)」の 3 段階で指定した。ほかの設定はベースライン最適化、パス保存なし、サムネイル保存なしとした。作成方法は、GIF の場合と同様に、TIFF データからの変換による作成を指定した。

(2) カラー画像

カラーについてもアンケートの回答をもとにしたが、PDF は画像データのみでの使用がそれほど一般的であるとは思えなかったため、検証対象とはしなかった。

①TIFF

選択理由はアンケートの結果による。ピクセル解像度の指定は、アンケートの結果にばらつきがあったため、当館の既存画像データの中では比較的オーソドックスな PhotoCD のフォーマットに倣った 1024x1536、2048x3072、4096x6144 ピクセル(それぞれ PhotoCD でいう 4base、16base、64base にあたる。それぞれ HDTV 出力、プリント出力、高品質プリント出力を想定した仕様)に決定した。ただし、古地図に関しては、このほかに「可能な限り高いピクセル解像度」を加えている。圧縮の有無は、アンケートでは二つに分かれていたが、可逆圧縮性を信頼して LZW 圧縮で指定した。

②JPEG

モノクロ画像と同様の理由による。ベースライン最適化、パス保存なし、サムネイル保存の設定も同様である。

③PhotoCD

当館が過去に作成したカラーの画像データは、ほとんどこの形式である。サンプルの作成は依頼しなかったが、既存のデータを使用して検証を行った。

④FlashPix

公的機関が策定した仕様に基づいたフォーマットではなく、また、開発から間もないためか実際のデータを目にする機会も少ないため、選択には議論の分かれるところであったが、ネットワーク利用を前提とした特徴が非常に魅力的であり、選択肢に加えることにした。

3.1.3 サンプル画像データ作成の手順

平成 10 年 10 月 2 日に、アンケート回答があった各企業に電子メールまたは FAX にて依頼状を発送。依頼状には検証内容、作成画像の種類、フィルムの種類などとともにフィルム選択の希望を確認する旨を記した。平成 10 年 10 月 9 日、22 日に承諾を

得た数社に対し、フィルム一式を郵送した。特定のフィルムを希望した企業には該当するフィルムのみを郵送した。

作成の締め切りは 10 月第 4 週までとしたが、フィルムの発送が遅れた企業もあり、サンプル画像データと元フィルムが揃ったのは 11 月中旬頃であった。

3.1.4 検証方法

(1) 体制

技術協力調整班メンバーおよびフィルム(=原資料)を所管する古典籍課、図書閲覧課の各課員が分担して検証に当たった。

(2) 使用ハードウェア

a) Power Macintosh (Mac 機)

型番:8100AV/100

CPU:PowerPC601 100MHz

メモリ 128MB

VIDEO:radius ThunderIV VRAM6MB

OS:漢字 Talk7.5.5

ディスプレイ:SUPERMAC 21T XL 1995.3 製造

ディスプレイの設定:1152x870、1670 万色表示

b) 日立 FLORA330DC2 (Windows 機)

型番:5DC02-YDOMA

CPU:MMX Pentium 166MHz

メモリ:64MB

VIDEO:ATI 3D RAGE LT VRAM2MB

OS:Windows95 ORS2.1

ディスプレイ:NANA O FlexScan E55D 1998.4 購入

ディスプレイの設定:1024x768、65536 色表示

当初は同じ画像データを上記の 2 機種で表示し、その違いを検証することを予定していたが、各設定のデータを抽出して表示を試みたところ、展開速度、判読性、精細度ともにほとんど違いが見られなかったため、どちらか一方の機種による検証を行うことに方針を変更した。

(3) 使用アプリケーション

- ・Imaging(Windows95 のリビジョンアップ版に付属のソフト、Windows 機で使用)

二値の TIFF の表示に用いた。

- ・Adobe Photoshop 4.01(両機種で使用)

二値以外の TIFF、JPEG、GIF、FlashPix の表示に用いた。FlashPix の表示に際してはプラグインを追加している。

JPEG と GIF の表示には WWW ブラウザを用いることも検討したが、表示倍率の変更がで

きないことから Photoshop を採用することになった。WWW ブラウザは FlashPix の特性を確認する際にのみ用いた。

(4) 具体的な検証方法

①精度の検証

作成データを特定の機器とアプリケーションを用いてディスプレイ上に表示し、目視にて比較検証した。原資料によっては、あらかじめ特定の文字や絵の一部などを検証のポイントとして指定しておいた。

②データサイズの検証

想定している用途(特にネットワークを通じた閲覧サービス)に適した精度のデータが、異なるフォーマットや設定において複数あった場合、データのサイズを比較して小さい方を選択した。データサイズは、Windows95 の MS-DOS 画面で dir コマンドを実行して記録した。ただし、PhotoCD データについては、拡張子 pcd のファイルサイズと、ipe ディレクトリ内にある 64base の差分ファイルサイズを別に記録し、それぞれ 1024x1536 と 2048x3072、4096x6144 ピクセルの検証結果と組み合わせた。

3.1.5 評価の項目と基準

評価項目は以下の9つを設定し、そのなかからデータに応じて評価項目を選択した。評価結果はすべての項目で3段階(+評価保留)とした。

(1)判読性 1

アプリケーションのウィンドウを最大化し(以下の項目も同様)、その中に対象資料の画像を隙間なく表示した状態(以下、全面表示と表記)で、対象資料に書かれた文字が読めるかどうかの検証。この場合、すべての画像において元データに対し約数%から50%程度の縮小表示となる。評価基準は、画像中に一部でも読めない文字がある場合(原資料自体に原因がある場合および検証者が当該文字を「文字」として理解できない場合を除く)は「1」、一通りの文字が読める場合には「2」、一見して明白に文字が読めると判断できる場合には「3」、とした。

(2)判読性 2

ソフトウェアの表示設定を用いて、ディスプレイの解像度に画像の解像度を一致させて表示した状態(以下、100%表示と表記)で、対象資料に書かれた文字が読めるかどうかの検証。評価基準は「a.判読性 1」と同じ。

(3)判読性 3

ソフトウェアの表示設定を用いて、100%以上の倍率で表示した状態(以下、拡大表示)で、対象資料に書かれた文字が読めるかどうかの検証。具体的な倍率は、文字を判読するための表示と考え、予め指定をしていない。評価基準は「a.判読性 1」と同じ。

(4) 精細度 1

100%表示で、画像中にある絵が鮮明に見えるかどうかの検証。評価基準は、画像中の絵の一部または全体が不鮮明に見える場合(原資料自体に原因がある場合を除く)は「1」、絵の全体が鮮明に見える場合は「2」、一見して明白に絵が鮮明に見える場合は「3」、とした。

(5) 精細度 2

拡大表示で、画像中にある絵が鮮明に見えるかどうかの検証。評価基準は、「d. 精細度 1」と同じ。

(6) 速度 1

アプリケーションのメニュー「開く」で画像データを選択し、「OK」をクリックしてから、画像が完全に表示されるまでの速度(以下、展開速度と表記)の検証。評価基準は、表示に 30 秒以上かかる場合が「1」、6～29 秒が「2」、0～5 秒が「3」、とした。評価基準の分けには特に客観的な根拠はなく、「1」が「許容範囲外」で、「2」と「3」が「許容範囲」とする検証者の見解の一致に基づいている。

(7) 速度 2

100%表示時または拡大表示時において、ウィンドウの右側と下側に出現するスクロールバーや Photoshop の「手のひらツール」機能などを用いて表示の範囲を変更する際の、体感的な速度(以下、スクロール速度と表記)の検証。評価基準は、引っかかるような、または這うような動きのスクロールが「1」、滑るような動きのスクロールが「2」、動態視力が追従できないほどの高速なスクロールが「3」、とした。

(8) 画質

モノクロ画像データの表示における、画像の質感に関する評価。特に基準を定めたわけではなく、評価の有無は検証者に任せた。実際に画質の評価については、具体的な状態を注記として書き込むこととした。

(9) 色再現

カラー画像データの表示における、画像の色彩に関する評価。カラーマッチングについては、技術協力調整班メンバーや館内関係者にエキスパートがおらず、今回の検証には保留の立場で臨んだが、原資料や同形式の他データと色彩が極端に異なる場合のみ記入できるように項目だけ設けた。何らかの評価結果が記入されていても、集計時には単なる参考データとして扱った。

3.2 評価結果の組み合わせによる評価

3.1 の評価結果を組み合わせることにより、以下に示すさまざまな角度からの評価を

行った。評価対象とした画像データの全件数は 1612 件である。

(注：1612 件は FlashPix と PhotoCD に関して異なる解像度で開いて検証している分を各 1 件と数えているため、実際のファイル数とは異なる。また、既定の検証には盛り込んでいない。また、協力企業が独自に作成した補正比較ファイルも含めていない。)

3.2.1 フォーマット+解像度による判読性の違い

同一の解像度においてフォーマットの違いによる判読性の差はほとんどなく、活字資料のモノクロフィルムで多値が二値よりやや劣り、木版および写本資料のモノクロフィルムで二値が多値よりやや劣る傾向が見られた程度であった。カラーフィルムにおけるピクセル解像度ごとの判読性も同様に差は確認できなかった。また、GIF フォーマットは全体的に他のフォーマットより判読性が劣るようである。

3.2.2 フォーマット+データサイズの違いによるファイルを開く速度の違い

同一解像度においてはデータサイズと表示およびスクロールの速度はほぼ比例し、フォーマットの違いによる明確な差は、GIF フォーマットがやや表示に時間がかかる以外に大差はなかった。

3.2.3 モノクロネガフィルムにおける 16mm と 35mm の相違点、カラーポジフィルムにおけるブローニーと 4x5 の相違点

16mm と 35mm との比較は、幕末新聞(木版刷り資料)を撮影した b 群と w 群において行った。結論としては、フィルム面積が大きい 35mm フィルムの方が判読性で劣ることが判明したが、その理由についてはさらに調査が必要である。

同様に、本草資料を撮影したブローニーと 4x5 についても比較してみたが、原資料の大きさがブローニー(縦 27cm)、4x5(17cm)とフィルムサイズに比例していなかったせいか、判読性と精細度で評価 2 以上とする比率がほぼ同じであった。

3.2.4 画像フォーマットごとの特徴

すべての資料群において TIFF を元に他フォーマットの画像を作成したためか、同じ元フィルムにおいて判読性や精細度の評価で TIFF より優れた他フォーマットは見られなかった。

(1)TIFF(二値)

活版印刷資料では概ね良好な判読性を示したが、印刷面に濃淡のある木版刷り資料や文字自体に濃淡のある写本の本草資料では、文字がかすれて判読不能な部分が生じる場合があった。データサイズが圧倒的に軽いだけあって、展開速度、スクロール速度は極めて優れていた。

(2)TIFF(多値)

他の設定が同じでも判読性にむらがあり、総合的な評価は迷うところである。全面

表示で許容範囲以上の評価であっても、100%表示では文字が滲んだように太く見えて、読みにくい場合が多かった。これは入力解像度の高低に関係なく、どちらかという会社ごとの特性と思われる。

(3) TIFF(カラー)

判読性、精細度は入力時のピクセル解像度設定に依存し、これといった特徴は見られなかった。データサイズは群を抜いて大きく、展開速度に時間がかかるため、一般的な閲覧の用途には不向きと思われる。

(4) GIF(今回の検証はモノクロのみ)

二値、多値ともに全面表示では全く文字が読めなかった。100%表示だと判読性が確保されることが多く、検証に用いた Photoshop だけでなくブラウザでも表示できる点と併せて有効性は高いが、実際に 100%での表示ではスクロールを伴って文字列を追うことになり、閲覧としての融通性は格段に落ちる。

(5) JPEG

低画質の設定だと、多くの画像に滲んだようなノイズが発生し、本草や錦絵といった絵の資料のみならず、活版や木版などの文字資料でも判読の妨げとなっていた。圧縮率と実用上で問題になるノイズ発生との関係は、今回の検証のような低画質—標準—高画質の 3 段階では明らかにならず、さらに細かい段階に分けた検証が必要である。展開、スクロール速度は高画質設定でもそれほど問題がなかったが、1024x1536 のピクセル解像度で高画質におけるデータサイズは平均で 500KB ほどあり、一般家庭からのインターネット経由による閲覧には不適切である。

(6) PhotoCD(カラーのみ。当館作成のものを使用)

これまで当館で一般的に作成されていただけあって、特に目立った可も不可も見あたらなかった。ローカルディスクにデータをコピーして 4096x6144 のピクセル解像度で表示する場合には、拡張子 pcd のファイルだけでなく差分データをディレクトリ構造をそのままコピーしておかねばならない点が煩わしく、実際の利用ではクライアント機の CD-ROM ドライブからの読み込みか、CD-ROM チェンジャに格納することになるものと考えられる。

(7) FlashPix

追加プラグインで表示可能にした Photoshop4.0 を用いて検証したが、FlashPix 本来の特徴である多解像度の内包および指定解像度または範囲のデータのみのネットワーク転送による展開速度、スクロール速度の向上性を生かした検証は、ネットワークの設定が困難で行えなかった。ただし、スタンドアロンによるブラウザでの表示では、20MB のデータでも問題なく扱え、ネットワークを設定し直して再検証する価値は十分にあると思われる。

3.2.5 モノクロフィルムにおける入力解像度(dpi)ごとの特徴

入力解像度の高さに比例して、画像の精度も向上するとは一概に言えない、というのが現時点の結論である。例えば、16mm 昭和期の a 群で、すべての判読性が評価 2 以上のものを抽出すると、TIFF のデータ 9 件が残るが、その内訳は二値の 200dpi が 2 件(母数 7)、同 300dpi が 3 件(母数 7)、同 400dpi が 2 件(母数 7)、同 600dpi が 1 件(母数 5)、多値の 300dpi が 1 件(母数 6)となる。いかなる要因で解像度の高さが失われるのかは、今後の検証における課題となろう。

3.3 フィルムごとの評価

3.3.1 16 ミリモノクロネガフィルム

(1)b 群. 幕末明治新聞

全 219 件。全面表示における判読性でいえば多値の 400dpi が最も適切であるが、データサイズが TIFF にせよ JPEG にせよ(GIF は判読性により除外される)平均 1MB を超えるため、インターネット経由の利用には問題がある。スタンドアロン環境および構内 LAN 経由での展開速度が許容範囲とみなされる。

(2)m 群. 明治期資料

全 207 件。全面表示ではほとんど判読不可能(全件のうち 2 件のみ、二値 300dpi の TIFF と多値 400dpi の TIFF)であり、スクロールを用いた表示にならざるを得ない。階調は二値で 200~400dpi、データサイズと展開速度を考慮すると 200dpi が適切である。フォーマットはデータサイズが比較的小さい TIFF または GIF とする。

(3)a 群. 昭和期劣化資料(a 函)その 1

全 211 件。全面表示で評価 2 以上はそのうち 11 件、すべて TIFF で、1 件以外は二値であった。入力解像度は 200~600dpi と幅がある。

(4)r 群. 昭和期劣化資料(a 函)その 2

全 185 件。全面表示で評価 2 以上はそのうち 7 件、すべて TIFF で、1 件以外は二値であった。入力解像度は 200~600dpi と幅がある。a 群と比較すると、撮影時期と撮影業者は異なるものの、対象資料の種別が同じだったためか、ほぼ同様の評価となった。

3.3.2 35 ミリモノクロネガフィルム

(1)w 群. 幕末明治新聞ほか木版刷り資料

全 150 件。全面表示では評価 2 以上がなく、100%表示で評価 2 以上のあるものを見ると、TIFF の二値 400dpi データが、作成した 6 件のうち 6 件とも該当していることがわかった。その TIFF から作成した JPEG と GIF についても、低画質の JPEG 以外は概ね良好である。同じ木版刷りの資料を撮影している b 群では多値が適切となり結論が異なったため、原因をさらに調べる必要がある。

(2)h 群. 本草資料

全 184 件。全面表示と展開速度において評価 2 以上のものは 12 件あり、フォーマットや入力解像度といった設定で見ればほとんどばらばらであるが、多値 400dpi の JPEG データは、低、標準、高の圧縮率がすべて含まれているため、ノイズが発生する 100% 表示を想定しないのであれば、低画質が実用的である。なお、多値 400dpi の TIFF データは展開速度に問題があった。

3.3.3 ブローニー判カラーポジフィルム

k 群. 本草資料

全 107 件。フォーマットと解像度については、どの設定の組み合わせでも、全面表示の精細度と判読性においては良好な結果であった。展開速度を勘案すると 2048x3072 ピクセル解像度以上の TIFF データと 4096x6144 の JPEG データがほとんど除かれた。母数に対する比率が最も高い設定は 2048x3072 の JPEG 高画質と標準画質(ともに 8 件中 8 件)であった。

3.3.4 4x5 インチ判カラーポジフィルム

(1)s 群. 本草資料

全 91 件。k 群と同様に絵が主体となる資料につき、100%表示時の精細度と判読性、展開速度で評価が 2 以上となるものを抽出したところ、形式(圧縮率)、解像度ともに多様なデータが 45 件挙がった。この中で、母数に対する比率が高かったのは 2048x3072 ピクセル解像度の JPEG(低、標準、高の圧縮率すべて)と 1024x1536 の TIFF、1024x1536 と 2048x3072 の PhotoCD であった。

(2)n 群. 錦絵

全 135 件。s 群と同じフィルムサイズで絵が中心となる資料ではあるが、被写体の面積としては寸法比で倍くらいある。抽出条件を s 群と同じにした場合、1024x1536 ピクセル解像度では PhotoCD(母数 8 のうち 8)と TIFF(母数 8 のうち 6)、JPEG 高画質(母数 8 のうち 5)の評価が高かった。2048x3072 ピクセル解像度では、PhotoCD(8 件)と TIFF(6 件)、JPEG の各圧縮率(それぞれ 6 件)の評価が高かった。

3.3.5 8x10 インチ判カラーポジフィルム

c 群. 古地図

全 123 件。かなり見た目の異なる 2 種類の古地図が混在し、しかも一方は分割撮影でもう一方は非分割撮影である。しかし、フィルムに写っている部分の大きさはほぼ同じで、判読性検証のポイントとなる文字の大きさも同じくらいであることから、集計上は特に区別をしなかった。古地図閲覧者の傾向を鑑みて、評価 2 以上が必須と思われる展開速度、全面表示の精細度、100%表示の判読性と精細度、それにスクロール速度でフィルタをかけると、28 件に絞られた。内訳は PhotoCD の 2048x3072 ピクセル(母

数は 9) と FlashPix(母数は 6)の 1024x1536 が 1 件、FlashPix の 2048x3072 が 2 件、1024x1536 の TIFF が 2 件(母数は 6。以下母数すべて 6)、2048x3072 の JPEG 標準画質が 2 件、2048x3072 の JPEG 高画質が 4 件、2048x3072 の TIFF が 5 件である。また、各社の最大入力解像度で作成を依頼したファイルについては、最大のもので 12366x13000 ピクセル(トリミング処理済み)、241.7MB あった。フィルムの情報をほとんど入力していると思われるが、それでも最も小さい文字は判読できず、展開には 19 分 50 秒を費やした。

3.4 指定外の処理

今回の検証で使用したサンプル画像データには、以下に示すような作成依頼時に指定しなかった処理を加えたものがいくつかあり、その処理のないものと比較すると、評価に少なからず差が生じている。このような処理は作業の手間から考え、大量の画像データを一括作成する場合に適用できるかは不明だが、一つの可能性として考慮する必要がある。

3.4.1 トリミング

資料本体を大きく表示できるよう、資料の周囲に存在する余白部分を削除する処理。検証に用いた一部フィルムには、撮影時の仕様のため広大な余白を持つものがあり、処理を施した画像データは判読性や精細度が大幅に向上していた。方法としては、フィルム全面をスキャニングした後に該当範囲を切り抜く方法と、該当範囲のみをスキャニングする方法が考えられる。ただ、大量の画像データを一括処理する際、折り込みページの展開等によるイレギュラーなレイアウトの画像データは正常に処理できないことがあり、処理前の設定と処理後の検査は入念に行う必要がある。

3.4.2 補正処理

アンシャープマスクというフィルタ処理を施した画像データは、処理のないものと比較して輪郭が鮮明になり、判読性や精細度の向上に寄与していた。ただ、設定値を上げすぎると不自然な感じが目立つため、処理前に適切な値を確認しておく必要がある。トリミングとは異なり、値が決まれば大量の画像データを一括して処理することが可能と思われる。

3.4.3 画像回転

人物を描いた錦絵に代表される縦長の資料は、画像データも一般的な横長から縦長に変更して表示することが望ましい。補正処理と同様に、一括処理が可能と思われる。

3.5 プリント出力による評価

画像データはディスプレイ上の閲覧提供のみならず、プリント出力による提供も期待される。今回の検証では、TIFF データに限って任意にプリント出力を依頼した。TIFF データに限定した理由は、他形式データの変換元として高品質であるからである。用

紙は当館複写サービスで最も利用の多い A4 サイズとし、種類は指定しなかった。

評価はディスプレイ上の評価と同様に、目視によって判読性と精細度（カラープリントのみ）を確認した。集計は原資料、入力階調（モノクロ）、入力解像度、用紙の種類ごとにカテゴリー分けをしてさまざまな分析を試みた。しかし、企業毎のカラープリントで入力解像度と判読性、精細度が比例する傾向が見られた程度で、カラープリント全般とモノクロプリントにおいては、これらの条件での特徴や傾向は確認できなかった。また、元画像データのディスプレイ上での評価とプリントの評価についても明瞭な関連性は確認できなかった。

3.6 検証の問題点

3.6.1 サンプル画像データの仕様

検討段階から、画像データの作成と検証には非常な労力がかかると予想されていた。そこで各画像データ形式の設定値を最小限に絞ったのだが、カラー画像のピクセルサイズ、JPEG 形式の圧縮率は設定値そのものが大雑把になり、検証結果から適切な仕様を導くことが極めて困難であった。特に JPEG 圧縮率の場合、同じ仕様のファイルでありながら、企業によってサイズに 10 倍以上も開きが生じ、「元となる TIFF データの 1/n」という設定の方が混乱が少なかったのではと考えられる。

3.6.2 サンプル画像データの作成手順

元フィルムから複数のデュープフィルムを作成し、各社には全く同じ条件の下でデータ作成を依頼することが本来望ましい。データ形式及びその設定の純粋な特性を比較検証できるよう、各社固有の特性を浮き彫りにして差し引く必要があったからである。しかし、実際には準備期間の短さその他の要因でデュープフィルムの作成ができず、原資料の大きさ、書かれた文字の大きさが異なるフィルムが少し混ざり、正確な検証の妨げとなった。

3.6.3 検証の方法

現在の検証体制では、目視による評価を主体とせざるをえないが、判読性や精細度の評価決定は、ある程度の基準を設けたとしても最終的には検証者の主観に委ねられることになる。主観を排除するためには、同一の画像データを複数の者が検証し、評価が分かれるデータについては、さらに別の者を加えて検証をやり直す方法が考えられるが、そこまで行うには至らなかった。

また、判読性の検証では、文字の大きさから視認に必要な最低限のピクセルサイズを算出し、実際の画像データの仕様とつけ合わせる手法も試すことが必要である。

3.6.4 集計、取りまとめの方法

評価の値は表計算ソフトに入力したため、抽出や並べ替え、グラフ化は容易であった。ただ、資料ごと、フィルムごと、データの形式や設定ごとの組み合わせは無数にあるが、評価が各項目につき 3 段階しかないのも、それぞれの特性を明確にするため

グラフ化が無効となる場合が多かった。また、上記のような問題点が積み重なると、集計の段階で不可解な結論が導かれることもあった。具体的な基準を見極めるには、より細かな評価段階を設定し、評価・分析を行う必要がある。

4. 総括

4.1 現在の利用環境において適切な画像データの形式

電子図書館の蔵書は来館利用と非来館利用、スタンドアロン PC による利用とネットワーク利用など、複数の利用形態が想定されるが、全ての利用形態に単一のデータ形式で対応するのでは、各利用形態の特性を生かした提供は行えない。以下では、調査検証のまとめとして、具体的な利用形態と組み合わせた画像データの形式について、以下のものが適切であると判断した。

4.1.1 スタンドアロン PC

(1) 想定環境

来館利用者向けのサービスを対象とする。機器は可能な範囲でできるだけ高性能のものを使用し、データをローカルディスクまたは高速な転送が可能なリムーバブルドライブから読み出すことが望ましい。特に Windows マシンの場合にはメモリを 64MB より多く搭載しないとメモリ不足で表示できないことがある。ソフトウェアは、予め登録されたサムネイル画像や予め付与されたキーワード等で画像データを選択できる機能を持つものを用いる。

(2) データ形式

①幕末から明治初期の雑誌・新聞、本草資料(無彩色)

モノクロのロールフィルム形態であれば、他の無彩色木版刷り資料、写本全般にも適用される。300～400dpi の解像度で入力した多値の TIFF データとする。解像度は、文字の大きさによって、資料ごとに適した値を設定する。200dpi と 600dpi の解像度については、判読性にばらつきが見られるため、採用を見送ることにすることが望ましい(他の資料も同様)。

②明治期刊行図書、昭和前期刊行図書

モノクロのロールフィルム形態であれば、他の活版刷り資料全般にも適用される。300～400dpi の解像度で入力した二値の MMR 圧縮 TIFF データとする。解像度は、文字の判読性によって、資料ごとに適した値を設定する。なお、当該資料中の図版を含むページについては、今回検証は行わなかったものの、木版や写本に似た性質を持つことが推測されるため、別に多値の TIFF で入力する。また、インターネットにおいてブラウザで表示することを前提とするならば、二値の TIFF は GIF に、多値の TIFF は JPEG に変換して使用する。これは、二値と多値を GIF と JPEG に変換した場合、二値では GIF の方が、多値では JPEG の方がファイルサイズが小さくなる傾向があるためである。

③本草資料(彩色)、錦絵

カラーフィルム形態であれば、他の彩色資料にも適用される。3072x2048 程度のピクセルサイズで入力した LZW 圧縮の TIFF データとする。ピクセルサイズは元となるフィルムの縦横比に合わせる。例えば 4x5 サイズのフィルムであれば 3000x2400 ピクセルとなる。もう一つの候補として PhotoCD 形式を挙げられるが、元フィルムの縦横比に関係なく画像が作成される(元フィルムの比率と PhotoCD の画像比率に合わない部分は黒く表示される)こと、上位へのピクセルサイズ変更はサイズを指定して開き直す必要があること、縦長の資料に合わせて回転処理を施した画像データが作成できないことが問題として存在することに留意すべきである。

④古地図

資料の大きさ、書かれた文字の大きさにもよるが、検証で使った 1 辺が 2 メートル以上の古地図だと、6144x4096 ピクセルの解像度では判読性に問題が生じるため、さらに高いピクセル解像度で入力する必要がある。また、撮影をする前に最も小さい文字のサイズを確認し、そこから逆算して分割撮影の枚数と入力の解像度を定めることも必要である。形式は LZW 圧縮の TIFF データとするが、利用時には 6144x4096 ピクセル以下に分割しないと、読み込みが遅くて使いものにならない。分割する場合には、スムーズな画像の切り替えが可能なように、システムの設計及びインターフェースのデザインを十分に検討する必要がある。

4.1.2 構内 LAN

4.1.1 のスタンドアロン PC、4.1.3 のインターネット以外に、特に構内 LAN 環境を想定したデータが必要であるかどうかは議論の分れるところである。しかし、ここでは仮に必要だと仮定し、望ましいデータ形式を示す。

(1) 想定環境

今回は構内 LAN の環境下における使用を想定した検証を行っていない。LAN 環境はクライアント機の仕様だけでなく、サーバーの性能、物理ネットワークや使用するプロトコルの種類といった全体的な構成でパフォーマンスが決定し、検証時に環境の想定及び構築ができなかったためである。従ってここでは、スタンドアロン PC よりは低速だが、インターネットよりは高速なデータ転送が可能で、クライアント機はスタンドアロン PC と同程度のものを用いると仮定する。また、ソフトウェアは、ブラウザを基本としてプラグインまたはデータごとの関連づけで他の画像表示ソフトが起動するように設定しておく。

(2) データ形式

①幕末から明治初期の雑誌・新聞、本草資料(無彩色)

スタンドアロン PC で使用する、300~400dpi の解像度で入力した多値の TIFF デー

タを変換した標準圧縮の JPEG データとする。変換する理由は、多値の TIFF より標準圧縮の JPEG の方がファイルサイズが小さくなるためである。圧縮率はネットワークの転送速度を考慮しつつ、判読性を確保できる値を設定する。なお、画像データは数千ピクセルのサイズとなり、そのままブラウザで表示するには難があるので、クライアント PC で画像表示ソフトを起動する方法か、利用者の指定に応じたピクセルサイズ(実際にはパーセントや「全面」「拡大」といった分かりやすい言葉等で指定)にサーバー側で変更してブラウザで表示する方法をとる。

②明治期刊行図書、昭和前期刊行図書

スタンドアロン PC で使用するものと同じ、300～400dpi の解像度で入力した二値の TIFF データを使用する。二値の TIFF データは JPEG よりも高圧縮なので、あえて変換は行わない。表示はブラウザではなく、別の画像表示ソフトを使用する。

③本草資料(彩色)、錦絵

スタンドアロン PC で使用する 3072x2048 程度のピクセルサイズで入力した LZW 圧縮の TIFF データまたは PhotoCD データから変換した低圧縮の JPEG を使用する。圧縮率はネットワークの転送速度を考慮しつつ、精細度を確保できる値を設定する。表示は他の資料と同様の方式を用いてもよいが、別にピクセルサイズを縦横半分程度に減らした JPEG 画像データを作成しておき、必要に応じて切り換えて表示する方法も考えられる。

④古地図

スタンドアロン PC で使用する LZW 圧縮の TIFF データと同じか、同形式から変換した低圧縮の JPEG データを使用する。表示は資料の特性上、スタンドアロン PC と同じものを流用することになる。

4.1.3 インターネット

(1) 想定環境

館内に設置される PC とは異なり、不特定多数の利用者が利用するため、環境は多種多様であるが、画像データを利用する場合には、多くはブラウザを用いてアクセスする。すると画像データの形式もブラウザ本体かそのプラグインによって表示できるものに限られるが、極めて低速なネットワーク接続である点も加味すると、圧縮率やピクセルサイズもある程度の妥協をみるか、可能な範囲で複数の選択肢を用意する必要がある。

(2) データ形式

①幕末から明治初期の雑誌・新聞、本草資料(無彩色)

構内 LAN 用の JPEG データを、文字の判読が可能な最低ラインまで再圧縮し、ピクセルサイズも落としたものを使用する。検証においては判読性が合格ラインで高圧

縮 JPEG データの最低ファイルサイズが 274KB(300dpi)である。ピクセルサイズの減少は検証を行っていないため、ここでは示さない。

②明治期刊行図書、昭和前期刊行図書

構内 LAN 用の TIFF データを変換した高圧縮 JPEG データを使用する。検証においては 312KB(300dpi)が 100%表示で判読性が確保される最小のファイルサイズである。そもそも、元の TIFF データは JPEG データよりも圧倒的にファイルサイズが小さい(上記データの変換前 TIFF は 59KB)ので、ダウンロード後にブラウザ以外の画像表示ソフトを使用することを前提に、TIFF データそのものを選択肢に含めることも考慮すべきである。

③本草資料(彩色)、錦絵

構内 LAN 用の JPEG データを、文字の判読性と精細度が確保される最低ラインまで再圧縮し、ピクセルサイズも 1536x1024 程度にまで落としたものを使用する。検証においては判読性及び精細度が合格ラインとなった高圧縮 JPEG データの最低ファイルサイズが 188KB である。他に、画像データ選択の目安となるサムネイル画像データ、判読性や精細度は問題あるが標準的なディスプレイに収まるサイズの画像データ、ファイルサイズは大きいものの構内 LAN 用と同等の品質を持つ画像データを可能な限り選択肢に加える。FlashPix 形式はインターネットにおける検証ができなかったため、今回は見送ることにした。

④古地図

構内 LAN 用のデータをさらに細分割し、1536x1024 程度のピクセルサイズに変換した JPEG データを使用する。他にも可能な範囲で分割(ピクセルサイズ)の大きい画像データを用意することも考えられるが、彩色本草や錦絵と異なり各画像データの相対的な位置関係を示す情報を別に用意しなければならない関係上、あまり効果的であるとはいえない。少なくとも現時点では、一般的なブラウザと JPEG 画像データの組み合わせにおいて、これ以外に効果的な表示方法は見当たらないのが結論である。

4.2 将来の利用環境を想定して作成しておく画像データの形式

画像入力作業に費用と手間がかかることは言うまでもない。また、元となるフィルムも永久にその良好な状態が保たれる訳ではなく、将来の画像入力は保証されていない。そこで、利用者の需要や利用環境に適した形式の画像データを作成すると同時に、将来における利用環境の向上を期待した、より品質の高い画像データを作成、保存しておくことも重要である。利用時の仕様よりも高い解像度で TIFF データを作成することにより、モニタ解像度の向上に対応した高解像度の TIFF データ、ネットワーク転送速度の向上に対応した低圧縮 JPEG データの作成が、新規にフィルムから入力し直すよりも低コストで可能となる。

4.3 画像データの閲覧に適したソフトウェアの機能、仕様

検証に使用したソフトウェアは、単に画像データを表示する以外にもさまざまな機能を搭載しているが、本来の用途が画像データの表示ではないため、必要とする機能が存在しない、存在しても使い勝手がよくない等の短所も抱えていた。ここでは検証を通じて電子図書館のスタンドアロンPC、構内LANクライアントPCにおいて画像データを快適に利用できるための表示ソフトウェアの機能、仕様を挙げる。

①表示サイズを選択

全面表示、100%表示、ウィンドウの縦(横)幅と画像の縦(横)幅を合わせた表示、任意の倍率の表示を画像表示前及び画像表示後に選択可能な機能。

②スクロール

ウィンドウに収まりきらない表示の時、スクロールバーや矢印キーだけでなく、マウスのドラッグ操作によって全方向へスクロール可能とする機能。

③表示モードの変更

二値－グレースケールの表示変更を、画像データの書き換えなしで行う機能。

④連続表示

連番をファイル名に用いている画像データを、簡易な操作で順番(逆順もあり)に表示する機能。

⑤プレビュー表示

表示する画像を選択する際に、先に縮小イメージを表示する機能。表示の有無はオプションにて選択可能とする。

⑥読み込みインジケータ表示

瞬時に画像を表示できない、ファイルサイズの大きいデータを開く時、進行状況を視覚的に明示する機能。

⑦画像の複数同時表示

ウィンドウを多重化し、同時に複数の画像を表示可能とする機能。

参考:検証に用いたソフトウェアの機能

ソフトウェア名	対応データ形式※	①	②	③	④	⑤	⑥	⑦
Photoshop4	MMR 圧縮 TIFF 以外	×	○	△	×	○	○	○
Imaging	TIFF, GIF	○	○	○	×	×	×	×
Internet Explorer4.01	JPEG, GIF	×	×	×	×	×	×	×
Graphic Converter	全て	×	○	△	△	○	○	○

○…機能あり、△…機能あるが使いにくい、×…機能なし ※…検証したデータ形式のみ

4.4 その他

4.4.1 画像データ作成にあたっての他の留意事項

今回のサンプル画像作成では、作成要項の書き方もわかりにくかったせいか、ファイル名のつけ間違いがいくつか発見された。発注者側では、間違いの発生を抑える作成要項(仕様書)の作成が求められる。

4.4.2 検証の体制

今回の検証は既存フィルムからの画像入力という範囲に絞り、少ない人員と短い期間で有効な結論を導き出せるよう計画を立てたつもりであったが、作業量を過少に見積っていたこともあり、実際には消化不良的な部分が残ってしまった面も否定できない。

今後の検証は、現行の組織を前提とすると、事前調査等の準備段階を含めて 1 年くらいの期間とより多くの職員の参加が必要である。

5. 今後の課題

今回の検証の対象としなかったり、対象に含めながらも十分な検証ができなかった以下の事項については、来年度以降に引き続き検証することが望ましい。

5.1 ネットワーク環境での検証

館のネットワーク担当者の協力のもとに模擬的な電子図書館システムを構築する。構内 LAN については、複数のクライアントから同時アクセスを試みる等の様々な条件下で、インターネットについては、多くの利用者環境と想定されるダイヤルアップ接続を再現し、実際の転送速度をふまえた実用的な画像データ形式及び設定を検証する。

5.2 原資料からの入力に関する検証

フィルム撮影を伴わない、原資料から直接的な画像入力の方法を検証する。各種スキャナやデジタルカメラ等を使用し、それぞれの方法と適合する資料形態の組み合わせとその設定を見極め、いずれの入力方法にも適合しない資料形態については撮影その他の代替手段を模索する。

5.3 彩色資料における色彩の再現性の検証

原資料の代替物として画像データを利用者に提供する際、彩色資料の場合は原資料と同じ色調を再現することがきわめて重要な要件となる。今回の検証において、各社から画像データの色再現に関する様々なアドバイスを得たが、検証に用いた環境で正確な色再現は実現できなかった。今後はまず、機器とソフトウェア、画像データの色調管理構造(カラーマネジメントシステム)について基礎から整理し直し、検証に必要かつ十分足り得る環境を構築したうえで、改めて取り組む必要がある。また、プリントアウト時における色彩の再現性についての検証も必要である。

5.4 今後の評価等の体制に関する提言

(1)未解決の検証項目

今回の検証は、電子図書館推進委員会技術協力調整班の枠組みの中で、特に検証用機器等の環境を用意せずに実施した。また、少数の職員が通常業務を抱えながら試行錯誤を繰り返して進めたものであるが、調査の結果、今後の在り方に関し、多くの課題と問題点が明らかになってきた。今回の対象資料はフィルム撮影済みの図書館資料

に限定したが、それだけでも膨大な検証項目がリストアップされ、多くの項目については時間的な制約から未検証の課題として先送りせざるをえなかった。しかし、今後当館が電子図書館を実現する場合には、これらの問題は未解決のままで置くことはできない。また、フィルム撮影済みの画像データ以外に、印刷物自体からの画像データ化や、さらにはHTML、SGML、XML等のテキストデータの検証等の課題も残されている。

(2) 継続的な検証の必要性

電子図書館の「蔵書」を構築するに当たって、電子データは多くのフォーマット等から最適なものを選択し、データを作成する必要がある。電子化を行うその最初の瞬間から、多くの技術的な選択肢の中で適切なものを選択しなければ、たった一つのデータも作成できない。また、それらがばらばらに選択され、標準化されていなければ、システムが統合的に取り扱うことができず、仮にシステムを作成しても、利用者にとっては使いにくいものになる。そのため、技術的な検証の継続的实施や標準化への取り組みは避けて通ることはできない課題である。電子データの保存を当館が実施するに当たっても、メディア変換、フォーマット変換、マイグレーション等は、今回の画像データの検証と同様の技術的な検証が必要であり、そのためにも継続した検証、調査・研究が不可欠である。

(3) 検証環境の整備

当館が電子図書館を実施するに当たっては、データの検証に基づく適切な品質管理を行うことは、今後継続して、また当館の業務として行うことが求められる。当面、画像の検証については、検証環境の整備を行う必要がある。部品の組み替え等によって多様な検証環境を構築できる複数の検証専用機器、機能の有無や優劣に応じて使い分ける各種ソフトウェアの導入は最低限用意する必要がある。また、ファイルサイズが大きい画像データを効率よく扱うためのファイルサーバの導入は必須であろう。画像データの表示に関わる機器やソフトウェアの進歩は著しく、検証機器の更新やインターネットを経由する利用者の環境を想定するために、常に業界の動向を把握している必要がある。また、画像データ形式の標準化や普及率にも目を配り、必要に応じて再検証を継続して行うことも求められる。

(4) 組織上の対応

それらの要件を総合的に判断すると、今後電子図書館を実現する上で、当館の施設・設備面でコンテンツの技術的な検証のできる環境を整えること、また、組織機構上にそれらを可能にする部門を設けることが必要であると考えられる。現在鋭意検討の進められている組織・機構の検討、関西館実施計画、電子図書館基盤システム計画のそれぞれの検討の中で、電子データの技術的検証（調査・研究、標準化の推進、検証・評価）等の部門の設置についてしかるべき配慮が求められる。また、組織再編以前においても、今後関西館（仮称）開館に向けた電子化を行うに当たり、電子図書館推進室が電子化データ作成部局と協力・調整の上、継続して電子データの検証等に当たる

ことが必要である。

なお、本報告書では、具体的な規模、組織名称等は明示的に示すことは行わないが、現在の組織機構の検討案（平成 11 年 2 月 経営改革検討委員会）に則つると、関西館に設置する「電子図書館部」、「図書館総合研究所」（いずれも仮称）等の組織に若干名の専任職員を配置するとともに、電子データの調査・研究を行う部署を設置し、必要に応じた外部専門家の招聘、関係機関との連携・協力事業を行うことが適当である。

<用語集>

CCITT : Comite Consultatif International de Telegraphique et Telephonique : 国際電信電話諮問委員会、現 ITU-T (International Telecommunications Union, Telecommunication Standardization Sector : 国際電気通信連合電気通信標準化部門。ITU の下部組織)

Flash-Pix : Kodak 社が Hewlett-Packard 社、LivePicture 社、Microsoft 社と協力して開発したファイル形式。一つのファイルに、取り込んだ画像の最大解像度から面積で 4 分の 1 ずつ小さくなった複数の解像度で画像を保持する。FlashPix 技術をサポートするアプリケーションでサイズの大きい高解像度ファイルを高速に転送及び表示できる。

GIF : アメリカのパソコン通信ネットワーク CompuServe で画像交換用に開発された画像形式。JPEG と共にインターネット標準の画像形式。CMYK モードはなく 256 色までしかサポートされていない。可逆圧縮によりファイル容量を小さくできる。ネットワーク用に開発されたフォーマットのため、画質より転送速度を早くすることを重視しており、インターネットで使用する画像に標準として採用されている。

G4 : Group 4。FAX の国際通信規格の一つ。アナログ回線を利用する G3 に対し、G4 はデジタル伝送路を利用する。

ICC : International Color Consortium。

JPEG : 静止画像データの圧縮方式の一つ。ISO により設置された専門家組織の名称がそのまま使われている。圧縮の際に若干の画質劣化を許容する(一部のデータを切り捨てる)方式と、まったく劣化のない方式を選ぶことができ、許容する場合はどの程度劣化させるかを指定することができる。方式によりばらつきはあるが、圧縮率はおおむね 1/10～1/100 程度。写真などの自然画の圧縮には効果的である。フルカラーで、印刷で利用するインクの色に合わせた CMYK モードをサポートする。

LZW 圧縮 : Lempel Ziv Welch。LZ 法を Terry Welch が改良したデータ圧縮方式で、画像ファイルの GIF やモデムの V.42bis など使われている。UNISYS 社の特許。

MMR : ITU-T T.6 に規定されている、G4 FAX 用の圧縮符号化方式。

PDF : Portable Document Format。Adobe Systems 社が開発した、PostScript ベースのドキュメントファイル形式。装置やアプリケーションに依存せずにドキュメント体裁をそのまま維持した電子配信が可能。オンライン配布向けに最適なファイルサイズに圧縮することができる。専用のソフト (Acrobat Reader 無償配布) が必要。

PhotoCD： Eastman Kodak 社と Philips 社が開発。通常のカメラで撮影したネガ／ポジフィルムを、紙焼きする代わりに、ISO 標準の CD-ROM にデジタル画像として記録する。媒体は Orange Book で規定された CD-WO (Write Once) を用い、ディスク 1 枚あたり最大 100 枚の画像データを収録することができる。

sRGB： RGB データの規格で、各種ハードウェアとソフトウェアのメーカーで保証されており、大半のスキヤナやプリンタ、アプリケーションの初期設定のカラースペースとしてよく使われる。RGB (Red / Green / Blue)はモニターに表示されている最も一般的なプロジェクトカラーの定義方式で、レッド／グリーン／ブルーのコンビネーションによる。

TIFF： Aldus 社と Microsoft 社によって開発された画像データのフォーマット。1 枚の画像データを、解像度や色数、符号化方式の異なるいろいろな形式で一つのファイルにまとめて格納できるため、比較的アプリケーションに依存しない画像フォーマットである。フルカラーで、印刷で利用するインクの色に合わせた CMYK モードをサポートする。

二値変換： 白黒二階調の画像に変換すること。

可逆圧縮： 画像伸長して元の画像に戻すときに元の画像の品位に戻る圧縮形式。(非可逆圧縮はその逆)

グレースケール変換： モノクロの階調のある画像に変換すること。

デュープフィルム： 複製フィルム。

シャープネス： 画像の鮮明度やピントの度合い。

モアレ： moire。格子模様の透過する物体（レースのカーテンなど）が重なったときなどに、本来の模様とは異なる大きな周期構造をもつ模様が観察される現象のこと。このように 2 つ以上の規則的な模様を重ねる場合などに見える、大きな周期の模様を「モアレ縞」と言う。印刷された画像（印刷時の網模様がある）を、イメージスキヤナなどで読み取る場合に発生し、画質を低下させることがある。

<参考文献>

- ・「画像フォーマットの基礎知識」：<http://www.bekkoame.ne.jp/~firefly/gform.html>
- ・「'98 最新略語辞典」：<http://home.impress.co.jp/%5Cmagazine/dosvpr/ryakugo/>
- ・「ASCII グローサリーヘルプ」：<http://www.ascii.co.jp/ghelp/>
- ・「画像の電子化とそのファイルフォーマット（特集 デジタル化の技術）」：加藤 茂夫、鈴木 馨 『情報の科学と技術』47(1) 1997.1 p9～16
- ・「目的別の画像ファイルフォーマット」：須藤 昇 『写真工業』54(10) 1996.10 p105～112
- ・「技術解説 画像ファイルの新方式 パソコン上の処理に最適一業界標準ねらう FlashPix の仕様と特徴一」：羽部 正義 『日経コンピュータ』439 1998.3.16 p187～192
- ・『Adobe Photoshop 5.0 Windows/Macintosh 版 ユーザガイド』 アドビシステムズ株式会社 1998.5