

メディアセンターにおける 150 年記念事業 — 記念写真と福澤著作コレクションデジタル化 —

いりえ しん
入江 伸

(メディアセンター本部調査役)

1 はじめに

メディアセンター本部デジタル担当では、学術情報のデジタルアーカイブの調査、開発、実験、運用のため、以下のようなプロジェクトを行っている。

- ・国立情報学研究所が進める次世代コンテンツ基盤構築委託事業(以下、CSI 事業)の委託を受けて進めているリポジトリ構築や電子図書館の再生事業
- ・慶應義塾大学学事振興資金を受けて取り組んでいる、博物館、文書館、図書館におけるアーカイブの相互運用性についての研究
- ・デジタルアーカイブリサーチセンターと進めている貴重書デジタル化と eRareBook の研究
- ・創立 150 年記念事業と連動したデジタルライブラリープロジェクトとしてのデジタルアーカイブ構築事業

本稿では、創立 150 年記念写真集の刊行事業の支援を目的に進めた慶應義塾写真デジタルアーカイブ構築と、CSI 事業の一部である電子図書館の再生及び、福澤研究センターとメディアセンターの共同プロジェクトとして取り組んだ、福澤諭吉著作全 119 冊の E-book 化作業の経過に沿って、デジタルアーカイブ構築に必要な、デジタイジング技術、コンテンツ知識、アーカイブ技術について説明する。

2 創立 150 年記念事業におけるデジタルアーカイブ構築の意味

創立 150 年事業でデジタルアーカイブ構築を行うことは、大きく以下の 3 つの意味をもつ。

- ・この機会に過去の資料を収集し、番号・分類・解説などのメタデータを付与し、組織化できる
- ・収集・整理した資料を長期的に保存できる可能性
- ・デジタルアーカイブを構築することで、インターネットの世界へ情報発信できる

創立 150 年記念事業を契機に、通常では難しい慶

應義塾の誇る歴史的な資料の収集を行い、これらを組織化し、長期的に保存し、Web 上に公開して誰でもアクセスできるようにすることは、歴史のある大学として文化的事業を先導することにもなり、評価を高めることになるだろう。

このような機会に資料を意識的に収集しないと、写真のネガやポジはロッカーに忘れ去られ、変色し利用できなくなってしまう。実際、今回収集された資料は、ネガや保存のための紙焼きではなく、アルバム、展示用のパネル、写真集が多かった。写真集を刊行するため写真を収集し、刊行後には掲載した写真を返却するか廃棄されてしまっていることが多いようだ。今回の作業では、収集した写真を印刷可能なデジタルデータとして保存し、素材アーカイブとして活用することができるようにした。また、今回収集した写真の中には、デジタルカメラで撮影されたものもあった。デジタルカメラが普及し、紙焼きが少なくなってしまうと写真の保存はさらに難しくなる。こうした状況を見ると、図書館が培ってきた収集、整理、保存の技術と新しいデジタル化技術とを駆使して、文化、社会への貢献をすることが求められていると感じる。

3 創立 150 年写真集出版事業支援アーカイブ

創立 150 年記念事業の一環である写真集の編纂のために、メディアセンターを中心として写真集編纂委員会が組織され、この事業の計画が立てられた。

これまでの写真集編纂は、収集された写真の複製を作るところから始まり、それを使って編集作業が進められている。例えば、福澤研究センターが業務用として活用している写真アルバムは、創立 125 年記念慶應義塾史の編纂のため作成されたと言われており、その資料価値は高く、今でも業務で使われるため、最近では写真集をすべてデジタル化し、パソコン上のアーカイブとして構築している。

今回の写真集編纂作業では、編集作業の前に収集

した写真をスキャニングし、印刷可能な精度のデジタルデータをつくり、それを必要なファイル形式へ加工し編集作業を支援し、印刷用データとして使用できるようにした。写真集の編集後には、著作権(複製権、肖像権等)を考慮したうえで、写真デジタルアーカイブとして公開する計画となっている。

(1) 慶應義塾写真データベースの反省

1995年にWeb公開し、全文検索機能を持った慶應義塾写真データベース^{注1}は、当時としては先端的なアーカイブであった。しかし、技術的には以下のような問題があった。

a 写真データフォーマットの問題

写真をスキャニングしたデータは、Kodakが開発したPhotoCDフォーマットで保存していたため、10年ほど前にKodakがPhotoCDのメンテナンス中止を発表したことにより、すべての画像を他のフォーマットへ変換しなければならなくなった。

b 特殊な全文データベースの問題

メーカーの特殊な全文データベースを使用し、その上に、特殊な機能をカスタマイズしたために、通常のメンテナンスが不可能となり、これも一般的なデータベースへ移行しなくてはならなくなった。データ移行をする際に特殊な形式を使ってツリー構造として表現していた写真の件名がデータベースから取り出せず、新しいデータベースではこの件名を捨てなくてはならなかった。

c 図書館内の体制

開発当時は技術の過度期ではあったが、データフォーマットやデータ構造が特定の仕様に依存しないように、技術標準を理解することや、データの相互交換、長期的運用についての理解が重要であった。

上記のような経験により、アーカイブ構築にあたっては、業者に頼るだけでなく、図書館側に十分な技術的な蓄積と体制が必要であることが分かる。

(2) アーカイブ構築のいくつかの道

デジタルアーカイブの構築には、いくつかの道筋がある。一つ目は新聞データベースや電子ジャーナル、今回の写真アーカイブのように、編集・出版工程、業務支援という明確な利用目的があり、そのためにデジタル化して、その副産物としてアーカイブを構築する道。二つ目は、図書館などで行っているように、これまで収集された資料を保存し、それを

Webで公開することを目的としてアーカイブを構築する道。三つ目は、最近活発に構築されている研究所内での研究活動の成果の蓄積として位置づけられ、成果の蓄積や共有、コミュニティー活動の中核としてアーカイブを構築する道である。

アーカイブ構築でもっとも難しいのは、一時的な流行に惑わされず、資料を収集する仕組みをつくり、アーカイブの目的を明確にし広報を行いながら継続的な運用体制を維持することである。そのために、デジタルアーカイブは長期的な組織や予算が確保できる組織が運用を担う必要がある。

10年前は、複製が可能な高精細な画像を撮影し、アーカイブしていくために、特性の高額な機材や特殊な技術を必要とした。また、技術水準がまだ貧弱であったため、民生機を超えた高精細画像でなければデジタルアーカイブではないという強い主張が確立されていた。しかし、最近の技術進歩によって、10年前の高精細の技術水準は、民生機レベルでも対応できるようになっており、機材のほかに撮影する資料へのストレスや撮影環境、撮影技術、撮影者という要素がデジタルデータ品質の分岐点になってきている。また、デジタルデータの最終的な品質は、資料の取り扱い方法、撮影方法、画像加工、ビューア、バックアップなどの連続した工程の結果として表れるため、工程での品質設計が重要となる。

高精細の複製を作るための品質と、文字が見えればよいという品質のレベルは、異なるものであり、それぞれの目的によって設計される必要がある。目的を曖昧にして、高精細なデジタル画像をつくろうとすると、ディスクもコンピュータもスキャナも、操作時間も、バックアップ媒体もすべてのコストが大幅に上がるために、継続的に運用できなくなってしまふ。それよりも予算や時間、人数によって、目的を達成するための品質設計をし、バランスがよく、安定した目的の品質を作り上げる技術が重要になっている。

(3) 写真アーカイブ構築までの準備

a デジタイジング仕様の決定まで

創立150年記念事業としての写真アーカイブは、最終的には写真集の印刷データになるため、それに耐えられるデジタルデータが必要であった。

印刷に耐えられるデジタイジングとはどういうことなのだろうか。どのようなことを仕様としてまと

める必要があるのだろうか。この点について簡単に解説する。

(a) 解像度

プリンタやスキャナなどで使われる解像度の単位は dot per inch (DPI) で表す。通常、写真集などの印刷では、出力線数(製版のアミの細かさ)の2倍の解像度となるので、175線のカラー製版出力の場合 $175 \times 2 = 350\text{dpi}$ が必要になる。35mmのポジをポストカードくらいの大きさに印刷する場合、35mmのフィルム面が、 $24\text{mm} \times 36\text{mm}$ 、ポストカードの印刷面が、 $102\text{mm} \times 152\text{mm}$ だとすると、ポストカードが4.25倍の大きさになり、印刷には、350dpi 必要なので、35mmのポジはその4.25倍の約1500dpi以上の精度が必要ということになる。もう一つの例として、単行本の撮影データをOCRにかけて利用しようとする場合、文庫本のサイズを、 $105\text{mm} \times 151\text{mm}$ とし、その一文字の大きさを $3\text{mm} \times 3\text{mm}$ とし、OCRをかけるためには、1文字32ドットが必要である。つまり、文庫本のページには、横 $105\text{mm} \times (40\text{ドット}/3\text{mm}) = 1400\text{ドット}$ 縦 $151\text{mm} \times (40\text{ドット}/3\text{mm}) = 2014\text{ドット}$ 画素数 $1400 \times 2014 = 2.8\text{M}$ 画素が必要となる。 $3\text{mm} \times 3\text{mm}$ の文字で40ドットで表現するためには、13ドット/mm、スキャナの300dpiは12ドット/mmとほぼ近い精度となる。通常は、文庫本よりも大きい面を撮影する上に、できるだけレンズの中心を使って撮影するため、18M画素くらいのデジタルカメラが必要となる。実際には、レンズ、撮像素子、カメラの画像処理の性能等が複雑に関係するために、画素数だけでは言い切れるわけではない。

(b) 色の表現力

コンピュータで表現される色はRGBという3つの要素で表現される。それぞれの色をどのくらいの階調によって、8ビット ($8\text{bit} \times 3\text{色} = 24\text{bit}$) と16ビット ($16\text{bit} \times 3\text{色} = 48\text{bit}$) が選択肢となる。ディスプレイやプリンタの進化も早いので、48bitが標準になっていくことを考慮すると、可能であれば48ビットでスキャニングすることが望ましい。しかし、48bitにすることによって、24bitよりも、ディスク容量が2倍になり、画像加工のソフトも限定されてしまうために、総合的な判断が必要となる。今回の写真アーカイブでは、48bitを採用した。

(c) 色空間とカラーマネジメント

カラー写真をスキャニングしてディスプレイに表示し、それをプリンタで出力したものを比べた場合、元の写真と同じ色になることはほとんど望めない。スキャナ、ディスプレイやプリンタは、再現できる色の範囲も機種によって異なるため、それを補正するための理論と技術をカラーマネジメントと言う。色を正しく処理しようとすれば、避けては通れない。今回のデジタル化にあたっては、それぞれハードウェアのカラープロファイルを作成し、データの相互交換時には、Adobe RGB色空間を使うことにした。

(d) スキャニング装置、パソコン等のハードウェア

ハードウェアの選定は、デジタル化する素材、目標品質、生産性、予算が決まった時点で最適なものを選定する必要がある。スキャニングする写真には、フィルムの場合は、ネガ・ポジの別、35mmや4×5などの大きさの別があり、紙焼きの場合は、カラーかモノクロの別、印画紙の大きさの種類、保存のされ方として、アルバムに貼られているか、台紙に貼られているか、保存状態はどのようになっているか等、いろいろな条件がある。紙焼きの大きさによって、A4のスキャナで行えるか、A2のスキャナが必要か、スキャニングする素材によって、フィルムスキャナが必要かなど、冗長さの高い機材を用意する必要がある。素材の状態が悪く、ガラス面へ密着させられなかったり、大きくてスキャナに載らない場合には、カメラを使って撮影する場合もある。

(e) フィルムや写真のクリーニング

古い写真の場合、写真が貼られた台紙が反っていて、スキャナにかからなかったり、カビが生えていて、影がでてしまったりする。スキャニングよりも保存を重視しなければならないことがアーカイブの感覚であれば、保存のためにどこまで被写体へストレスを与えていいかということは、常に重要な課題となる。このため、写真を技術として理解し経験のある専門家の助言が必須である。

(f) 担当者の適性

撮影をするフォトグラファーやスキャニングの担当者には、スキャニングする被写体への関心と興味が重要である。何気なく資料を扱う時や、長時間単調な作業を続ける時に、特にそのような要素が重要になってくると感じている。デジタイジング担当を

雇用する場合、専門教育を受けているか、被写体に興味と愛着を持っているかが重要な判断材料となる。

b 写真の「保存とデジタル化」のワークショップの開催

上記のように多様な選択肢を一気に確認するために、2006年7月27日に「写真デジタルアーカイブ/デジタル化と保存のためのワークショップ」を開催した。古写真を理解するために、日大芸術学部の高橋先生、写真のデジタル化では実績のある堀内カラー、美術関係の印刷業者に参加してもらい、総合的な議論を行った。その結果、紙焼きは600DPI、ネガ、ポジは、2000DPI、カラービットは48bitを基準とすることにした。

c メタデータ仕様

写真の場合は、図書館が主に扱う図書資料と異なり、そのタイトルも解説もほとんどない。そのため、資料としてアーカイブするためには、メタデータの付与が必須である。最低でも、その写真に写っている建物、物、人、撮影した日時、場所が必要である。こうした内容を記述するためのExcelの入力シートを新たに作成し、各部署へ写真収集を依頼すると同時に、メタデータの付与も依頼した。このメタデータを活用するためには、相互交換が可能な形式にして公開する必要があるため、慶應義塾大学学術情報アーカイブシステム（以下、KOARA）のArticleアイテムタイプからPhotoタイプを分離し、上記のExcelシートからKOARAのPhotoタイプへと変換を行った。

(4) 写真アーカイブの将来

今回は、写真集の編纂事業として行っているため、KOARAのプライベートなエリアで公開を制限している。これらの写真を公開していくには、公衆送信権の大学への許諾や、プロジェクトの継続と保存のための仕組みを確立していく必要がある。

写真は、近世の歴史的な資料として保存されている貴重な媒体である。これを整理し、公開していくことは、これからの教育、研究資料として重要な課題である。今回のプロジェクトがこれで終わるのではなく、継続的な事業として展開されている、デジタルカメラで撮影した画像データの収集・保存の検討も進めながら、継続していくことが重要である。

4 福澤著作コレクションアーカイブ

(1) 福澤研究センターとの共同事業

創立150年を記念したメディアセンターと福澤研究センターの共同事業として、福澤の著作55点(版本)全119冊をデジタル化して公開するプロジェクトを進めてきた^{註2}。1年間に渡るデジタイジング、Webデザイン、著作の解説に対する公衆送信権の許諾を受け、2007年7月に公開した。慶應義塾とGoogleの連携の記者会見においてもデモを行い、高い評価を得た。また、創立150年記念講演会の全会場においてもポスターを掲示し、慶應義塾の貴重なコレクションを紹介することになった。

(2) 貴重書デジタイジング技術

貴重書デジタル技術と設備は、慶應義塾大学HUMIプロジェクトで確立された技術移転を受け、その技術的コアをベースにコストを下げて量産化するためのノウハウと設備の整備を行った。

HUMIプロジェクトでは、グーテンベルグ聖書を所蔵する図書館へ遠征し、そこで確実に、安全かつ短期間に最新技術を使った高精細画像を撮影するため、大規模な技術、手法、設備を用いていたが、それによって作業に多くの冗長性が生まれてもいた。しかし、慶應の図書館内での撮影は、自館が所蔵している資料であるため、それぞれの冗長性をなくし、撮影方法をシンプルにすることが可能となる。また、図書館で撮影するものは、和装本が多く、インクナブラのような本を100度しか開けないという制約もなく、ストレスにも耐久性が高いため省力化が可能となっている。また、技術の進歩によって、デジタルカメラも最高技術を求めるというよりも、安定して操作性が高いものを選択しても十分な精度が確保できる。

一方で、フィルムで撮影し、それをスキャニングしてデジタル化する方法を取っているところもあり、その有効性もある。しかし、その場合フィルムの保存という問題や、工程が増えることによるコストアップがあり、本プロジェクトでは、現資料を保存することを前提としたデジタル撮影の方針を取った。

(3) 学術的な解説

図書館がWebで公開する際には、メタデータだけではなく、その資料に関する学術的な解説が重要となる。特に、本プロジェクトでは、公開にあつ

て、重要な特徴として位置づけている。福澤著作コレクションの公開に当たって、各著作の解説には、富田正文氏のご遺族からの許諾をいただき、『福澤論吉書誌』(大塚巧芸社、1964年)を、各巻表示の解説には、佐志傳先生と雄松堂の許諾を受け、『マイクロフィルム版福澤関係文書。収録文書目録第2分冊・福沢論吉関係資料(1)』(慶應義塾福澤研究センター編・雄松堂出版、改訂再版1998年)を引用することにし、その特性を生かして見ることができるようWebデザインを検討した。

(4) ビューアーの利用

デジタルコンテンツは、ビューアー上パッケージされた表現である。ビューアーによって、そのコンテンツの評価さえ変わってしまうこともある。福澤著作コレクションでは、ロゴスウエア社製のFLIPPERという、閲覧しやすいページめくり式のインターフェースを使っている。このビューアーは、2006年1月の慶應義塾図書館貴重書展示会で、貴重な資料のため、ガラスケースに入れられ、他のページを見ることができないというフラストレーションを解消するために、利用して高い評価をうけ、それ以降、標準的なビューアーとして採用している。

このプロジェクトでもっともコストがかかったのは、撮影以降の工程であるビューアーへのパッケージ化するための画像加工である。撮影した画像を1ページずつ、切り出し、影を消す作業は、手間のかかる作業である。また、撮影抜けやパッケージ化の抜けなどのチェック作業もコストがかかる作業である。これからは、パッケージ化の効率化、作業工程のチェックを工夫していく必要がある。

(5) デジタルコンテンツの活用

デジタル化によって、インターネット上で利用できるようになった貴重なデジタルコンテンツは、工夫次第で様々な活用できる。インターネット上の多様な知識とリンクし、融合されていくことで、研究、教育、ビジネスのスタイルを大きく変えていくこと

も可能である。このような可能性を追求しながら、貴重なデジタルコンテンツを財産として保存していくという一見矛盾したことを進めていく必要がある。そのためにも、デジタル化の技術的な課題解決を委託に任せるのではなく、独自で評価し解決することが必要だと思う。デジタル化の技術进行评估し、デジタルデータを活用していくには、内部での柔軟で高品質なデジタル化ラインが必須だと考えている。

5 おわりに

デジタルアーカイブを構築するためには、アーカイブ資料の理解、デジタル化技術への理解、メタデータや相互運用性に対する理解、WebやデータベースなどのIT技術へ対する理解、デザイン力、コンテンツの活用方法に対する理解などの総合的な力が必要である。また、インターネットの世界では、図書館の資料も博物館の資料も、美術館の資料も、アーカイブの資料も、研究所の成果も、地域資料も、全てのコンテンツが相互に利用される対象となり、知識として融合されながら使われようとしている。これからの図書館にとって、多様な組織との連携が重要であり、そのような関係抜きにはデジタルアーカイブの存在は語れない。これからの多様なデジタルアーカイブの連携へ、メディアセンターがなんらかの役割を果たせていければ、150年記念の事業としても意味のあるものになっていくと思う。

最後になったが、このアーカイブの完成のために協力していただいた皆さんに感謝したい。

注

- 1) 慶應義塾写真データベース。(オンライン), 入手先<<http://photodb.mita.lib.keio.ac.jp/>>, (参照 2007-08-06).
- 2) 慶應義塾メディアセンターデジタルギャラリー。(オンライン), 入手先<http://project.lib.keio.ac.jp/dg_kul/>, (参照 2007-08-06).