

第3章 資料・記録のデジタルデータ化について

震災関連の資料・記録には、アナログ情報とデジタルデータがあります。アナログ情報をデジタルアーカイブに保存・公開するためには、デジタルデータ化作業が必要です。

第3章は、震災関連の資料・記録の収集からデジタルデータ化し、アーカイブで公開するための準備として必要な作業について説明するものです。「3. 震災関連デジタルアーカイブでの事例」では、運用実証調査の事例を紹介しています。

1. 資料・記録の収集について

(1) 震災関連の資料・記録

震災関連の資料・記録には、アナログ情報とデジタルデータがあります。アナログ情報をデジタルアーカイブに保存・公開するためには、デジタルデータ化作業が必要です。

なお、対象と考えられる資料・記録は、災害の記録だけではなく、被災前後の画像や避難所、仮設住宅での生活の様子を伝える記録、避難当時やその後の体験談、支援や救助の記録、被災前の地域の行事等の記録など、多種多様です。

どのようなアーカイブを作るかによって、収集する資料・記録は異なりますが、震災関連の記録や記憶を伝える資料が散逸しないように、まず収集しておく姿勢が大切です。

特に、書籍・刊行物は、図書館等において、ジャンル・分量共に豊富な資料・記録を収集していますが、被災地のみで配布されている広報紙や個人の手記等、地元の機関、団体だからこそ収集できる非常に貴重な資料もあります。

資料・記録の収集に際しては、震災関連の資料・記録を所有する他の機関（省庁等、地方自治体、都道府県立図書館、放送（テレビ／ラジオ）、新聞、NPO・ボランティア団体、学会・研究機関、大学、海外、その他）に協力してもらう方法も考えられます。東日本大震災関連の資料・記録を保有する機関と主な資料・記録について「参考4 東日本大震災被災関連情報の収集先」にまとめました。

運用実証調査で収集したコンテンツをもとに、下記に震災関連の資料・記録と収集先機関を整理しました。アナログ情報には、表3-1に示したように、文書、画像、音声、動画その他の資料があります。運用実証調査では、個人、自治体や地元企業、支援活動を行ったNPOなどから幅広くコンテンツを収集しました。

表 3-1 アナログ情報の種類と例

種類	アナログ情報の例
文書 ¹⁸	震災関連の記録集、震災に関する調査報告書、論文集、復興に関する計画書、救助・復旧活動等の計画書、活動記録集、新聞記事、避難所だより、壁新聞、メルマガ、震災関連のイベント・セミナー・相談会等のチラシ・配布資料、個人・団体が作成した手記、体験記録、文集、震災に関する内容のフリーペーパー、ミニコミ誌、チラシ、学校だより（同窓会・PTA）、会報、広報誌、個人が発行した手記、詩集等（自費出版物も含む）

¹⁸ 文字情報のみでなく、一部に画像情報等を含む場合もある。

種類	アナログ情報の例
	商工会会員の被災事業者数等のデータ 等
写真、画像	震災関連の写真集、活動記録集（画像）、報告書（画像）、新聞（写真）、震災前・震災後の画像、空撮写真等
音声（カセットテープ等）	儀式での式辞、あいさつ、祝辞、送辞、答辞、誓いのことば、オーラルヒストリーに係るヒアリング記録、コミュニティ FM 等
映像、動画（ビデオテープ等）	録画等映像資料（視聴覚資料も含む）、語り部の記憶を伝えるインタビュー映像等
その他	地図、気象図等

資料・記録をデジタルデータ化して保有している機関も多数あります。表 3-2 に、デジタルデータを保有している主な機関とその特徴を取りまとめましたが、ほかにも、被災された方や支援を行った個人の方、NPO ほかボランティア団体、企業や商工会議所、農協、漁協などの団体が保有している写真や動画データも多数あると考えられます。

運用実証調査では、様々な機関や個人の方にコンテンツの提供を依頼したほか、一般の方からの投稿を受け付けているポータルサイトとも連携し収集を行いました。

また、NDL 東日本大震災アーカイブは、多数の震災関連デジタルアーカイブと連携していますので、NDL 東日本大震災アーカイブと連携することにより、NDL 東日本大震災アーカイブが連携している他の震災関連デジタルアーカイブとの連携も可能です。

表 3-2 デジタルデータの種類と収集している機関、特徴など

デジタルデータの種類	収集している主な機関	主な特徴	主なデータ形式
テキスト情報（書籍含む）	省庁等 地方自治体 大学	各省庁・自治体は所管する分野の震災関連情報（被害状況、被災関連手続きの案内、防災関連情報、復興状況等）や調査レポート等が提供されている。	PDF
ウェブページ	行政機関 自治体	報道発表／プレスリリース、またテキスト情報と同様の震災関連情報（あるいはやや簡略化した情報）が提供されている。	html
学術データ	大学	震災関連の研究活動やプロジェクトに関する紹介・報告資料等が公開されている。	PDF
投書・手記	新聞 放送（テレビ）	震災関連のニュース・特集・読者・視聴者からのメッセージがまとめて公開されている。	html
ブログ、ソーシャルメディア（SNS、ツイッター等）	各種サービスプロバイダ（行政のツイッター含む）	震災関連の活動に参加している方が当事者の視点で、体験したこと感じたことを発信している。府省や自治体でもソーシャルメディアの利用が進んでおり、災害情報や避難所情報をソーシャルメディアが提供されている例もある。	html

デジタルデータの種類	収集している主な機関	主な特徴	主なデータ形式
写真、画像	省庁等 放送（テレビ） 新聞 大学 ポータルサイト	省庁等、報道機関、大学等が公開している画像には、震災時の被害の状況や救出・救援活動の様子に関する記録画像が多くみられる。一方、民間のポータルサイトでは写真共有の仕組みを用いて画像を公開しており、被災時のみならず被災前の画像とともに公開するなど、被災地の記憶を伝える役割も果たしている。	JPEG（様々な媒体からのアクセスを可能とするため、一般的な形式とすることが多い）
ラジオ	放送（ラジオ）	市区町村から発表されるライフライン・生活関連情報や、NPO や企業などの団体が行う支援活動、支援イベント等が提供されている。	MP3
音声	放送（テレビ）	被災地の方の証言記録のインタビュー音声を公開している。証言マップによる検索も可能。	MP3
映像、動画	放送（テレビ） ポータルサイト 地方自治体 自衛隊等	放送局が公開している動画には、被災者からのメッセージ・証言等の資料・記録が多くみられる。一方で、デジタルカメラや携帯電話の動画機能の普及により、一般の方が撮影した震災時の動画が大量に存在している。ポータルサイト等ではこのような個人から提供された動画の共有の仕組みが提供されている。地方自治体や自衛隊等で、活動記録等をDVD化している場合もある。	WebM MPEG4 3GPP MOV AVI MPEGPS WMV FLV 等

(2) 資料・記録の収集方法とポイント

震災関連の資料・記録は地方自治体や図書館だけでなく、企業、NPO、個人の方など、様々なところから集めることが可能です。できるだけ広く声をかけてみましょう。

資料・記録を提供していただく際の方法は、対象がアナログ情報かデジタルデータかによって大きく異なります。

運用実証調査で実施した資料・記録の収集手順は、以下のとおりとなります。

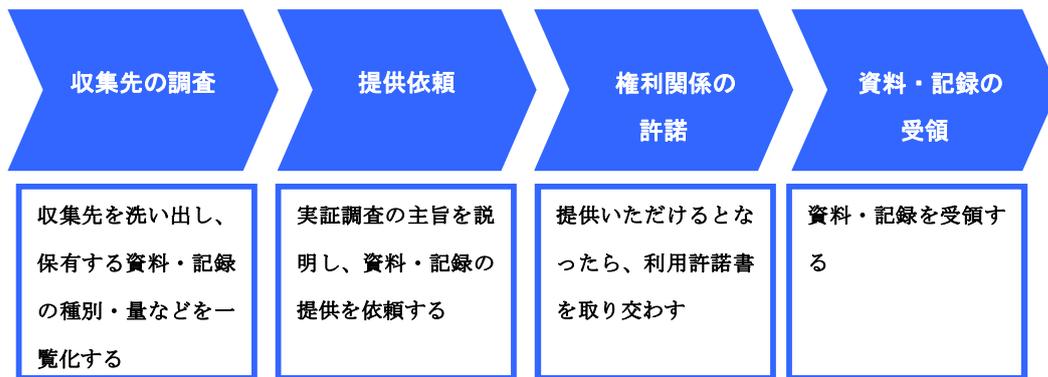


図 3-1 資料・記録の収集方法

(a) 収集先の調査

資料・記録の収集先について、連絡先や保有する資料・記録の種別、量などを一覧表にし、「資料・記録の収集リスト」を作成します。

(b) 提供依頼

資料・記録の保有者に対して訪問又は電話等で資料・記録収集の主旨、目的などを説明し、協力の可否の意思確認を行います。

(c) 権利関係の許諾

資料・記録保有者が資料・記録を提供する意思がある場合、利用許諾書を取り交わします。権利関係処理の詳細は、「第 6 章 震災関連デジタルアーカイブの権利関係」をご参照ください。

(d) 資料・記録の受領

資料・記録を受領します。デジタルデータの場合、後日、媒体で送付していただくか持参した媒体に複製してもらい等により受領しますが、アナログデータの場合は、デジタル化が必要となるため、借用してデジタル化した後、返却するか、持ち出しできない資料の場合は、スキャナを持ち込み、デジタル化を行うなどの方法で受領することとなります。

受領した資料は、目録を作って管理することをお勧めします。運用実証調査では、資料・記録の提供者がわかる名称、入手日、受け取った人、資料の種別（写真、動画、文書など）、二次利用に関する許諾の種類、資料内容に関する簡単な説明など、資料・記録を受け取った人以外でも作業が円滑にできるよう、資料・記録ごと、あるいは提供者ご

とに資料・記録を管理しました。デジタル化以降の作業の進捗を管理する上でも、このような工夫は有効です。特に、目録を作成しておく、メタデータの作成や付与の際にも役立ちます¹⁹。

2. デジタルデータ化のフロー

運用実証調査で実際に用いたデジタルデータ化の一般的な作業フローと、写真、紙媒体といったアナログ情報の種類（デジタル化の対象）ごとのデジタル化手法のポイントを以下に整理します。

なお、ここにあげた内容は、運用実証調査における手法をもとに記述したのですが、アーカイブの運用ポリシーや地域の実情に応じて、作業を簡略化しても良いでしょう。

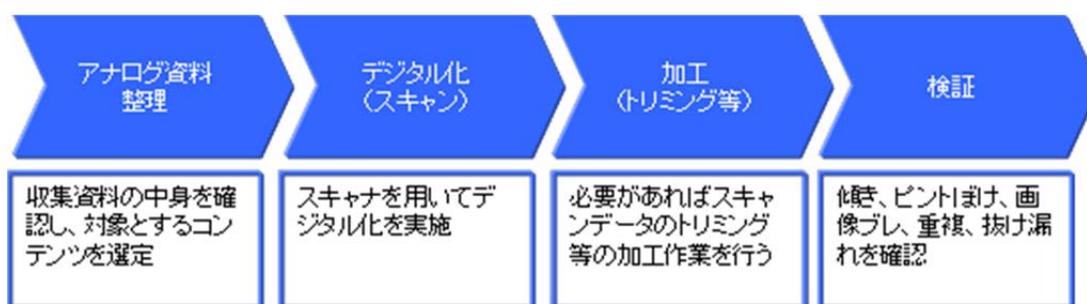


図 3-2 デジタル化の作業フロー

(1) アナログ資料の整理

デジタル化による利便性等の利点として以下のような点が挙げられます。

- ▶ デジタル化し、適切な保存措置をとることで、汚損した資料の代替として安定的な保存が可能となる。
- ▶ デジタル化した情報をクラウド等の環境に保存することでいつでもどこでも閲覧・利用が可能となる。
- ▶ 原本の代わりにデジタル化して提供することにより、原本をより良い状態のまま保存することが可能となる。

一方、デジタル化作業には人手や費用がかかるため、優先順位を決めておくことも重要です。

優先順位を決めるうえで重要なポイントは、構築しようとしている震災関連デジタルア

¹⁹ 本ガイドライン「第5章 コンテンツのメタデータ付与について」参照。

アーカイブの目的や目標を、運用ポリシー²⁰の一部として、まず初めに定めることです。対象とすべき資料・記録の中でデジタル化すべきアナログ情報の種類や特徴、必要な費用、デジタルデータの価値や利用の緊急性等は、アーカイブの目的や目標に照らして最終的に判断すべきものです。

デジタル化の対象とすべきアナログ情報の優先度を決定する際のポイントを、「セーブ・ザ・メモリープロジェクト」²¹の事例をもとに、以下に取りまとめます。

《優先度を検討する際の考慮事項》

① デジタル化対象の劣化状況

特に銀塩写真において写真表面が汚損された場合は、写真表面のゼラチン部分の細菌による腐食を防ぐため、優先的に洗浄・デジタル化を行う必要があります。

※すぐにデジタル化できない場合でも、細菌の腐食を抑えるため、汚損された写真を冷凍保存することをお勧めします。

② デジタル化の費用

デジタル化作業にどの程度の予算をかけられるかにより、全体としてデジタル化できる資料のボリュームが決まります。予算の範囲でどの資料を優先的にデジタル化するか、決める必要があります。

地方自治体と同プロジェクトの支援を受け、汚損写真の洗浄～デジタル化までの作業を行う体制を整備・運用した場合の費用の例を次に示します。

- 専任の臨時職員 2 名の人件費
- 設備費用（洗浄・デジタル化を行う場所）
- スキャナ（複合機）2 台（無料貸し出し、ただし故障対応のため、保守契約だけ締結）
- PC（安価なものを無償提供）
- サプライ代（写真のシリアルナンバーを印刷する A6 用紙、写真枚数に比例）
- インターネット回線（自治体と通信事業者が個別に契約）
- クラウド利用料（写真の保管用）

③ デジタル化に係る時間

デジタル化作業にどのくらいの期間をかけられるかにより、全体としてデジタル化できる資料のボリュームが決まります。予定している作業期間の範囲内で、どの資料を優先的にデジタル化するか、決める必要があります。

²⁰ 運用ポリシーについては、本ガイドラインの第 1 章を参照のこと。

²¹ 株式会社リコーに対し、ヒアリング調査を実施。「セーブ・ザ・メモリープロジェクト」については、本ガイドライン「第 2 章 被災資料の応急措置、修復、保存について」にも記載。

「セーブ・ザ・メモリープロジェクト」における各工程の大よその所要時間を以下に示します。

洗浄工程： 3段階で行い、合計5分程度。

乾燥工程： 陰干し・自然乾燥のため、湿度によって所要時間が大幅に異なる。(低湿度の日には2時間で乾くこともあるが、湿度が高いときには丸1日かかることもある。)ドライヤー等による乾燥は、写真がカールしてしまったり、伸ばした際にしわ・ひびが入ってしまうことがあるため導入していない。

スキャニング： 1日8時間で1000枚程度、慣れてくると1,200枚程度の生産性がある。(ただし、大規模に実施するのであれば設備化が必要となる。)

④ 費用対効果

デジタル化する費用とデジタル化することのメリットより、デジタル化しないことのデメリットを優先して判断するほうが、優先順位を決めやすいかもしれません。

上記はあくまでも「セーブ・ザ・メモリープロジェクト」の事例における目安です。機材によっては生産性が向上し、期間が短縮できたり、ボランティア等を募って費用を下げることであったりしますので、本格的な作業に着手する前に、試行作業を行ってみると良いでしょう。同プロジェクトでは、マニュアルを用いた作業の標準化、フラットベッドスキャナを用いたスキャニングの高速化、デジタルデータのアップロード用のアプリケーションの改善等で作業の生産性の向上を図っています。

試行作業を行うことで、その結果をもとにマニュアルを作成するなど、本格作業に向けた準備にもなるでしょう。

また、これらの要素から、画一的に優先度が定められるわけではなく、実際には個々の状況に応じた判断が必要となります。

(2) デジタル化

優先順位が決まったら、デジタル化作業を実施するための機材を準備し、ファイル形式を決めます。ファイル形式を決める際には、長期保存²²にも配慮しましょう。

(a) デジタルデータ化のための機材

アナログ情報をデジタル化するための機材としては、各種スキャナやデジタルカメラなどがあります。また、スキャナで読み取ることが難しい文字情報をキーボード入力することも考えられます。対象資料について事前に調査を実施し、その種類・特性・状態を判断した上で、適切なファイル形式及びそれに即したデジタル化の方法を選択することが重要です。紙媒体の図書館資料のデジタル化については、「国立国会図書館資料デジ

²² 長期保存の留意点については、本ガイドライン第4章を参照のこと。

タル化の手引 2011 年版²³」では、以下の 2 つの手法を紹介しています。

- ▶ 原資料をスキャナ又はデジタルカメラで直接スキャニング又は撮影する。(※)
- ▶ 原資料のフィルム撮影を行った上で、撮影したフィルムをフィルムスキャナ等でスキャニングする。

(※) 原資料を直接スキャニングする場合は、原資料の形態及び必要とする解像度に応じて、次表に示すスキャナ等の中から選択する。

表 3-3 デジタル化に用いるスキャナ²⁴等の種類と特徴

スキャナの種類	特徴	対象となるアナログ情報の種類
デジタルカメラ	原稿台上向きに対象資料を置き、原稿台から離れた上面から撮影が可能であるため、立体物や厚みのある資料のスキャニングに使用する（オーバーヘッドスキャナと同様）。	文書、写真、画像 ※直に触れてはいけない資料等
フラットベッドスキャナ	フラットベッドスキャナは、資料をスキャナの前稿台にセットして、原稿台ガラスの下に取り付けられた読み取りヘッドがデータを取り込む形式のスキャナである。原資料の厚みがなく破損の危険性が低い場合や、高解像度でスキャンする必要がある場合に使用する。資料を裁断することが可能である場合や一枚物の文書類のスキャニングの場合、オートマテックドキュメントフィーダを使用すると、連続してデジタル化することが可能である。ただし、厚みのある冊子体資料の場合は、ガラス面に強く押さえつけることにより破損する恐れがある。	文書、写真、画像 ※破損の危険性が低い資料等
オーバーヘッドスキャナ	オーバーヘッドスキャナは、原稿台上向きに対象資料を置き、原稿台から離れた上面から撮影する形式のスキャナである。原資料が厚みのある場合に使用する。原資料を強く押さえつけずにデジタル化できる。	文書、写真、画像 ※直に触れてはいけない資料等
その他	特殊な用途で使用するハンドヘルド型やシートフィーダ型がある。シートフィーダ型は、資料の裁断が可能である場合、連続してデジタル化することが可能である。	文書、写真、画像

出所：「国立国会図書館資料デジタル化の手引 2011 年版」を基に MRI において作成

デジタル化に使用する機会が多いと思われる機器について、各々の費用、メリット・デメリットに関する分析を以下に取りまとめます。

²³国立国会図書館「資料デジタル化の手引 2011 年版」

<http://www.ndl.go.jp/jp/aboutus/pdf/digitalguide2011.pdf>

²⁴ スキャナは、強い光を発する場合がある光に弱い資料（建築の青焼きや、像の薄くなった戦前期の写真など）には注意が必要である。

表 3-4 デジタル化手法ごとの費用、メリット・デメリット

機器	費用	メリット	デメリット
フラットベッドスキャナ	複合機等の大型機の場合、導入費用は大きいですが、リース等も可能。大量のデジタル化が必要な場合は生産性が高い。小型機の場合は、導入費用は小さい。	高速に処理できる。	接触するため対象を損傷させる可能性がある。
オーバーヘッドスキャナ、デジタルカメラ	普及品のデジタルカメラであれば、導入費用は小さい。ただし、生産性は低い。	接触させないため対象の損傷を防ぐことができる。	処理に時間がかかる。非接触であるため再現性に限界がある。
OCR	キーボードによる手入力に比して機器導入費用がかかる。特に複合機等の大型機の場合、導入費用は大きい。	高速に処理できる。	厚い冊子は認識が難しい。フラットベッド型の場合は、接触により対象を損傷させる可能性がある。
キーボード	導入費用は小さいが、生産性が低く人件費の負担が大きい。	接触させないため対象の損傷を防ぐことができる。文字であればテキスト内容をより正確にデジタル化できる。	処理に時間がかかる。

(b) データのファイル形式

▶ 国立国会図書館が推奨するフォーマット

「国立国会図書館資料デジタル化の手引 2011 年版」においては、画像データベースや保存用に使用する画像フォーマットとして比較的高品質・高圧縮な TIF 又は JPEG2000 を推奨しています。また、用途に応じて GIF、PNG、JPEG 又は PDF の画像フォーマットを選択しても良いでしょう。

NDL 東日本大震災アーカイブのように、誰でも利用できる公共性の高いシステムにおいては、ファイル形式の汎用性も重要な条件となります。「セーブ・ザ・メモリープロジェクト」においては、デジタル化した写真データをクラウドに保管し、写真の持ち主が携帯や PC など様々な環境から写真を閲覧できるようにするため、最も汎用的な形式として JPEG 形式へのデジタル化作業を行っています。

▶ 国立公文書館で採用しているフォーマット

参考として、国立公文書館で採用しているフォーマットをご紹介します。国立公文書館では、標準的フォーマット、長期保存フォーマット、デジタルアーカイブ用フォーマットの 3 種類のフォーマットを使っています。

表 3-5 データの種類別定義

種類	標準的フォーマット		長期保存フォーマット	デジタルアーカイブ用フォーマット
文書作成	OASYS 一太郎 8-12 Word 97-2003 Word 2007	PDF PDF/A OpenOffice Writer	PDF/A	PDF/A
表計算	Excel 97-2003 Excel 2007	OpenOffice Calc	PDF/A	PDF/A
プレゼンテーション	PowerPoint 97-2003 PowerPoint 2007	OpenOffice Impress	PDF/A	PDF/A
画像	JPEG JPEG 2000 GIF	TIFF BMP	JPEG2000 (lossless)	JPEG2000 (lossy) PDF (JPEG2000)
音声	WAVE MP3	WMA	MP3(ビットレート 256kpbs以上)	MP3(ビットレート 256kpbs以上)
映像	QuickTime Windows Media	RealPlayer MPEG	MPEG-2	MPEG-2

出所:

<http://www8.cao.go.jp/koubuniinkai/iinkaisai/2011/20110908/20110908haifu6-1.pdf>

まず、一般的な文書については、OASYS、一太郎、Word といった市販のソフトウェアによるもののほか、PDF、PDF/A、さらにオープンソースである Open Office によるファイルも受け入れています。一方、長期保存のためには、それらのファイル形式を PDF/A 形式に統一しています。

同様に、表計算、プレゼンテーションについても長期保存のためには、PDF/A 形式としています。PDF/A 形式に変換しておくことで、特定の商用アプリケーションがないと利用できないといった依存性を持たせずにデジタルデータを利用することができます。

一方、画像については、JPEG/JPEG2000/GIF/TIFF/BMP といったフォーマットを、JPEG2000 に変換します。音声については、MP3 に、映像については、MPEG-2 にそれぞれ変換して保存します²⁵。

➤ 運用実証調査で使ったフォーマット

運用実証調査においては、「知のデジタルアーカイブ社会の知識インフラの拡充に向けて」（総務省 2012年3月）、「国立国会図書館資料デジタル化の手引き 2011年版」等を参考に、フラットベッドスキャナを用いたスキャニング方式を採用しました。解像度は、印刷して視認可能であることを基準としています。

²⁵ 音声、映像で採用している MP3、MPEG-2 等は、保存容量を重視した非可逆圧縮であり、画像で採用している標準化を意識した JPEG2000 とは、事情が異なると思われる。

表 3-6 運用実証調査におけるデジタル化作業の概要

地域	デジタル化対象	デジタル化精度 (視認可能)	ファイル形式	速度	作業員数	読取の阻害要因
青森	行政文書	・ 文書 : 200dpi ・ 画像 : 300dpi	PDF JPG	40 枚/時間	1 名	付箋やホチキス留め 折り込んである資料
岩手	紙資料	・ 文書 : 300dpi ・ 画像 : 300dpi	PDF	75 枚/時間	管理者 1 名 作業員 5 名	ページ数の多い冊子はスキャン負荷が莫大
宮城 (東北大学)	図書館刊行物	・ 文書 : 300dpi ・ 画像 : 400dpi	PDF JPG	0.5~1 件/時間	2 名	特になし
宮城 (河北新報社)	新聞記事 (イメージ)	・ 文書 : 500dpi	PDF	測定していないため、不明	7 名	OCR ソフトでの文字判読は困難 (※)
福島	広報誌・文書	・ 文書 : 300dpi ・ 画像 : 400dpi	PDF JPG	20 枚/時間	2 名	紙文書の付箋

※新聞記事は、形が不定形であることから OCR で読み取るとは困難であるとの実証調査結果となりました。

その他、具体的なデジタル化手法の詳細は、以下の資料に詳述されています。

◇総務省：知のデジタルアーカイブ「デジタルアーカイブの構築・連携のためのガイドライン」参考資料 9 斯道文庫「デジタル化の基礎知識」、2012 年 3 月

http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000041.html

◇国立国会図書館：資料デジタル化の手引 2011 年版、2011/8

<http://www.ndl.go.jp/jp/aboutus/digitalguide.html>

- ◇国立国会図書館：平成 24 年度資料デジタル化研修（実践編）講義資料、2012/10
http://www.ndl.go.jp/jp/library/training/material/1196986_1486.html

- ◇日本写真学会：文化財写真の保存に関するガイドライン、2012/5
www.spsti.org/item/pdf/1337264652_event_pdf2.pdf

- ◇東京大学大学院情報学環・凸版印刷：
文化資源のデジタル化に関するハンドブック、2011/11
<http://www.center.iii.u-tokyo.ac.jp/handbook>

- ◇東京都歴史文化財団 東京文化発信プロジェクト室：
アート・アーカイブ ガイドブック β 版、2011/3
http://www.art-society.com/parchive/new/guidebook_beta.html

- ◇歴史資料保全活動におけるデジタルカメラによる文書資料撮影の手引き
<http://www.miyagi-shiryonet.org/01/satuei04/satueihou04.htm>

- ◇樫村 雅章：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、2010/05

(3) 加工

デジタル化作業が完了したら、アーカイブで公開することを考慮し、必要に応じてマスキングやぼかしなどの処理を行います。

運用実証調査では、以下の作業を行いました。

➤ 個人情報のマスキング

「公開可能」として許諾いただいた資料・記録であっても個人が特定される情報があった場合には、マスキング処理を実施しました。

➤ 画像方向の修正

おもに写真で表示の向き（縦・横）を修正しなければならないものがありました。

➤ 音声データの分割

複数ファイルが 1 コンテンツになっている場合があり、分割処理が必要な場合がありました。

➤ Web サイトアーカイブ時のトリミング

Web サイトは、画像（JPG 形式）としてキャプチャし保存しました。広告部分等余計な個所は削除対象としました。

(4) 検証

全ての作業が終わったら、解像度の設定が正しいか、判読可能かなどの確認作業を行います。せっかくデジタル化した資料が、後々使えない、ということにならないよう、この段階で確認し、必要に応じて画像ブレや傾き、ピントぼけの修正作業を行います。また、同じ資料・記録を重複してデジタル化していないかの確認をしておくことも重要です。

運用実証調査では、以下の観点で検証作業を行いました。

▶ 解像度の設定が正しいか

画像は鮮明なほうが見栄えが良いが、高精細すぎるとファイル容量が大きくなりシステムにかかる負荷が大きい。また、画面表示に時間がかかるため、表示速度と画像の鮮明さとのバランスを考えた設定とする。

▶ 判読可能か

画面で文字が読めるか、印刷したときに読めるかを確認する。

▶ 台帳の記載と登録画面が一致しているか

収集したコンテンツの目録にある資料・記録のうち、デジタルデータ化することとした資料・記録が、重複なく全てデジタルデータ化されているかを確認する。

3. 震災関連デジタルアーカイブでの事例

運用実証調査では、震災関連資料・記録を収集し、そのうち、デジタルデータ化されていない紙資料等のアナログ情報についてデジタルデータ化を実施しました。

本項では、(1)において、資料・記録の収集とアナログ情報のデジタルデータ化の手順について、運用実証調査全体での作業概要を説明し、(2)以降では、各運用実証調査での事例を説明します。

(1) 運用実証調査でのデジタルデータ化

(a) 資料・記録の収集

① 資料・記録収集結果

運用実証調査では、公共機関を中心に、民間企業、学術機関、個人、NPO 法人等の団体からデジタル及びアナログ情報を収集しました。各地域の収集先別に集計した資料・記録数については、以下のとおりとなります。

表 3-7 資料・記録収集結果

地域		公共機関	民間企業	学術機関	個人、その他団体	計
青森	収集先	府省、自治体	地元企業、インフラ企業、地元メディア、商工会議所	八戸工業大学、その他県内の大学	個人、医療法人、介護福祉施設	
	コンテンツ数	30,772	9,500	数十	-	40,272
	比率	76%	24%	0%	-	100%
岩手	収集先	自治体、図書館	地元企業、地域メディア、青年会議所	岩手大学	個人、ボランティア団体、NPO	
	コンテンツ数	1000	8,820	1,240	5,245	16,305
	比率	6%	54%	8%	32%	100%
宮城 (東北大学)	収集先	大学付属図書館	-	東北大学	個人、研究者	
	コンテンツ数	989	-	82,559	17,186	100,734
	比率	1%	-	82%	17%	100%
宮城 (河北新報社)	訪問先	-	地元メディア	-	個人、ボランティア団体、NPO	
	コンテンツ数	-	51,324	-	22,796	74,120
	比率	-	70%	-	30%	100%
福島	訪問先	府省、自治体	インフラ企業、地元メディア	-	個人(ボランティア団体)、医療機関	
	コンテンツ数	30,362	2,252	-	82	32,696
	比率	93%	7%	-	0%	100%

また、各地域で収集した資料・記録は、文書、写真、映像等、様々なものがありました

た。資料・記録の種別で集計した資料・記録数については、以下のとおりとなります。

表 3-8 資料・記録収集結果

地域		写真	音声	映像	文章	計
青森	詳細	災害関連写真	-	災害関連映像、復旧作業映像	行政決済文書、震災関連刊行物、学術レポート	
	コンテンツ数	33,500	-	-	6,772	40,272
	比率	73%	-	-	17%	100%
岩手	詳細	災害関連写真	FM 放送記録等	被災記録等	新聞記事・取材資料、オンライン記事	
	コンテンツ数	12,139	201	74	3,891	16,305
	比率	74%	1%	1%	24%	100%
宮城 (東北大学)	詳細	災害関連写真	オーラルヒストリー	-	学術論文、チラシ、パンフレット、古書	
	コンテンツ数	98,918	776	-	1,040	100,734
	比率	98%	1%	-	1%	100%
宮城 (河北新報社)	詳細	災害関連写真	-	-	新聞記事見出し	
	コンテンツ数	23,194	-	-	50,924	74,120
	比率	31%	-	-	69%	100%
福島	詳細	災害関連写真	FM 放送記録	インターネット放送	各自治体 HP、震災関連刊行物、プレスリリース情報、Twitter データ	
	コンテンツ数	17,936	350	72	14,338	32,696
	比率	55%	1%	0%	44%	100%

② 資料・記録収集の作業体制

資料・記録収集においては、収集に要した期間により人員数に変動がありましたが、各プロジェクトともおおむね 2~5 名の作業体制で実施しました。

なお、資料・記録収集作業では、資料・記録提供者に対して本事業の主旨を説明し、

了解を得る必要があったため、収集作業にあたる人員は、プロジェクトの目的や構築するシステムについて、理解していることが必要となります。

以下に各地域の作業体制を示します。

表 3-9 資料・記録収集の作業体制

地域	収集期間	役割	人材	スキル・トレーニングの有無	人員
青森 資料・記録総数： 40,272	2012/10 ～ 2013/3	市民、民間企業資料・記録収集者	地元企業職員	無	5名
		行政文書収集者	開発業者社員	無	3名
岩手 資料・記録総数： 16,305	2012/10 ～ 2013/2	収集スタッフ	NPO、企業、団体のスタッフ	地元の地理、情報を知っていて、地域団体とのネットワークを有するのが望ましい	2～3名
宮城（東北大） 資料・記録総数： 100,732	2012/11 ～ 2013/1	事務局、収集担当者	大学生、地元住民	・方言への対応や、土地勘を有する人材が望ましい ・インタビューのトレーニング、機器の説明が必要	事務局：2名、収集担当者：16名
宮城（河北新報社） 資料・記録総数： 74,120	2012/11 ～ 2013/2	-	-	-	-
福島 資料・記録総数： 32,695	2012/9 ～ 2013/3	収集担当者	-	学識経験者、アーカイブ、システムの知見を有する人材が必要	2～3名

(b) デジタルデータ化について

① デジタルデータ化作業結果

運用実証調査においては、受領した資料・記録のうち、アナログ情報については、デジタルデータ化作業を実施しました。

各地域のデジタルデータ化作業の概要と作業効率について、以下に示します。

表 3-10 デジタルデータ化作業概要

地域	デジタルデータ化対象	資料・記録量	読取の阻害要因	読取作業効率
青森	行政文書	約 3,000 件	付箋やホチキス留め、折り込んである資料	40 枚/時間
岩手	紙資料	約 900 件	大規模の冊子ではスキャン負荷が莫大となる	75 枚/時間
宮城 (東北大学)	図書館刊行物	約 1000 件	特になし	1 冊当たり 2 時間程度
宮城 (河北新報社)	新聞記事 (マイクロフィルム)	約 58 万ページ	OCR で読取れない場合がある	1 記事当たり 1~2 時間程度
福島	プレリリース資料	約 250 件	紙文書の付箋	20 枚/時間

② デジタルデータ化のファイル形式

運用実証調査において、アナログ情報をデジタルデータ化する場合の最適なファイル形式や精度をプロジェクト毎に決めました。

基本的には、「国立国会図書館 資料デジタル化の手引き 2011 年版」の①保存用画像の項において「原資料に対して 300~400dpi」との記載があり、これに準じて定めることとしましたが、利活用の際の利便性を考慮し、本運用実証調査では、下記の設定で問題がないことを確認しました。

- 文書では 200dpi で十分確認可能（目視で確認して概ね問題が無い）。
- 画像に関しては研究利用を想定した場合に最大限の解像度が望ましいものの、現実的には 300dpi 程度が妥当であると思われる。

表 3-11 ファイル形式 (再掲)

地域	デジタルデータ化対象	デジタル化精度 (視認可能)	ファイル形式	備考
青森	行政文書	・ 文書：200dpi ・ 画像：300dpi	PDF JPEG	-
岩手	紙資料	・ 文書：300dpi ・ 画像：300dpi	PDF JPEG	画像のうち写真は600dpiとした
宮城 (東北大学)	図書館刊行物	・ 文書：300dpi ・ 画像：400dpi	PDF JPEG	-
宮城 (河北新報社)	新聞記事 (マイクロフィルム)	・ 文書：500dpi	PDF	画像データの他、テキストデータを作成
福島	広報誌・文書	・ 文書：300dpi ・ 画像：400dpi	PDF JPEG	-

③ デジタルデータ化作業体制

デジタルデータ化作業体制は、資料・記録の内容や量によって大きく異なるため、作業体制は、柔軟に対応できるようにしておく必要がありました。

運用実証調査では、紙文書のデジタルデータ化であれば、作業には特別なスキルが必要とならないことから、雇用創出の観点から積極的に地元の人材を登用しました。

ただし、写真や映像等の中には、震災の記憶をよみがえらせてしまうものもあったため、心情的な配慮が必要となりました。また、上記の人員で、スキャナ等の機器の操作や不具合への対処のため、機器の操作に精通した管理者を体制に設置することが必要です。

以下に各地域の作業体制を示します。

表 3-12 デジタルデータ化作業体制

地域	役割	人数	必要スキル
青森	・ 作業者 デジタルデータ化	1名 市庁 OB	スキャナの操作そのものの習得についても時間を要しなかった
岩手	・ 作業者 デジタルデータ化	5名	事前の研修により、未経験者も可能であった
	・ 管理者 問合せ対応	1名	機器の操作に精通した要員を確保した

地域	役割	人数	必要スキル
宮城 (東北大学)	・ 作業者 デジタルデータ化	2名	特になし
宮城 (河北新報社)	・ 作業者 デジタルデータ化	7名	特になし
福島	・ 作業者 デジタルデータ化	2名	スキャナの操作の習得には時間を要するが、スキャン作業担当者は既にスキャナ操作は習得済みであったため、操作方法の習得に時間は要しなかった

(2) 青森県における実証調査

青森プロジェクトでは、青森県が被災県であると同時に支援県でもあるため、資料提供元の所在地により、被災関連資料と支援関連資料の両方が収集できることを期待し、資料・記録の収集を行いました。また、震災後早い時期に復旧・復興段階へ進んでいることから、震災直後から復旧段階、復興段階の資料を網羅的に収集できることが期待されるため、以下の観点で資料・記録収集を行いました。



- ・ 同一箇所、同一テーマに関して、震災前～復興までを時系列で比較検討
- ・ 防災対策の震災前の想定と、実際の被害を比較検討
- ・ 復興計画のモデルとしての情報を収集

デジタルデータ化が必要となった資料・記録は、行政関連資料のうち、紙で未整理の状態でも保管されていた資料となります。

ここでは、八戸市で実施した行政資料のデジタルデータ化作業について紹介します。

(a) デジタルデータ化の手法・プロセス

① デジタルデータ化の手順

八戸市では、未整理の行政資料をデジタルデータ化することとなりました。

資料は市庁舎外への持ち出しが禁止となっていたため、市庁舎内に機器等を持ち込み、デジタルデータ化作業を実施することとなりました。

作業者は、八戸市元職員 1 名を雇用し、未整理の資料を整理することから実施しました。

はじめに、資料の目録を作成し、八戸市職員がデジタルデータ化の要／不要を判断しました。スキャン作業は、目録単位で実施しました。

② デジタルデータ化に使用した機器等

デジタルデータ化で使用した機器等は、以下のとおりとなります。

表 3-13 デジタルデータ化における機器等

機器	台数	備考
EPSON DS-7000	1	A3 フラットベッド/シートフィード (1 分当たり 70 枚)

資料はホチキス留めや付箋の貼られているものも多かったため、ADF（自動紙送り装置）を使用できず、1 枚 1 枚スキャンすることとしました。そのため、1 分当たり 0.7～1 枚程度の読み取り速度でした。

(b) デジタルデータ化における課題と対応策

デジタルデータ化における課題と対応策については、以下のとおりです。

① スキャン作業の作業負荷

【課題内容と対応策】

八戸市庁における行政文書のスキャニングに関しては、ADF 付のスキヤナを持ち込んで作業を行いました。付箋や写真の貼付のある資料や、折り込んである資料は結果的に 1 枚 1 枚をスキャンする必要があり、非常に時間がかかることが分かりました。また、ホチキス留めの文書が多く、ホチキスを外してスキャンした後の原状復帰作業にも時間がかかり、デジタルデータ化の効率が向上しませんでした。

有効な対応策が無かったため、作業人員を増やすことで作業時間の長期化を防ぐこととしました。スキャン作業は 1 枚当たり 1 分から 1 分半程度の時間を要するため、1 時間で 40 枚程度の作業効率でした。

② スキャン作業の場所について

【課題内容と対応策】

資料・記録の持ち出しについて許諾が取れない場合があります。

八戸市庁に関しては、資料・記録の持ち出しができず、八戸市庁内に作業場所を確保してスキャン作業を行うこととしました。

(参考) 八戸市作業風景



③ ファイル分割単位の決定

【課題内容と対応策】

スキャン単位、資料・記録単位を事前にルールとして定義して作業をスタートしましたが、行政文書の内容を理解していなければ「決裁単位」が判断できない資料があり、資料・記録収集先への確認を必要としました。

(参考) 資料の整理状況



八戸市との協議で、ファイルの最小単位を1決裁文書又は、添付資料などの参照資料がある場合には、資料の種類ごとにファイルを作成しました。

例) 決裁文書に3つの添付ファイルがある場合には、決裁文書を含め、4つの資料・記録となります。

④ デジタルデータ化資料・記録のタイトル名の付与について

【課題内容と対応策】

スキャン後、デジタルデータを保存する際、資料・記録の内容から判断してタイトルをつけることは、高度なスキルが要求されることから、タイトルとしては、スキャンした時のファイル名をそのまま使用する形で作業を実施しました。

ところが、実際にシステムで使用するにあたっては、資料・記録の内容が想像できないファイル名では、メタデータ付与等の作業で使いにくいことが判明し、デジタルデータのファイル名について、メタデータの項目から自動生成することとしました。

その際、フリーソフトを使用し、ファイル名の一括変更を実施しました。

(3) 岩手県における実証調査

岩手プロジェクトでは、資料・記録収集における方針として、「震災資料の収集先の体系を調査し、多様なセクターから実証的に収集し、アーカイブモデルの構築を行う」ことを掲げました。具体的には、収集の現場としては陸前高田市、大槌町を中心に、収集対象に関しては、特定の種類の団体、機関からの資料・記録収集に偏らず、企業、メディア、地方自治体、NPO等、広く資料・記録収集を行いました。

岩手プロジェクトにおいて、デジタルデータ化が必要な資料・記録は「おおつちさいがいエフエム」の紙資料（番組進行表）900点のみでした。ここでは、その作業概要について紹介します。

(a) デジタルデータ化の手法・プロセス

① デジタルデータ化の手順

紙資料は、以下の選定基準により、デジタルデータ化するものと不要なものを判断しました。

表 3-14 デジタルデータ化の選定基準

選定基準	選定基準詳細
プロジェクトの目的と合致していること	震災前、震災直後、復興のプロセスと時系列的に幅広い分布の資料であるかどうか。
文書量が確保されていること	1つの団体でまとまった数量の資料を保有していること。
メタデータ付与のための情報があること	メタデータ付与のための情報がある程度整備されており、問合せ可能な人員が収集先に存在していること。
「貴重」と思われる資料であること	震災直後の記録、記憶が散逸しやすい資料など、他には存在しない、貴重な資料であること。

② デジタルデータ化に使用した機器等

デジタルデータ化に使用した機器等は、以下のとおりとなります。

表 3-15 デジタルデータ化における機器等

分類	メーカー名	型番 等	備考
画像スキャニング (単票)	Panasonic	KV-S4065CWN	自動原稿送り機能 A3 サイズまで対応 標準附属ソフト利用
画像スキャニング (冊子等)	plustek	Opticbook A300	フラットベッド型 A3 サイズまで対応 標準附属ソフト利用

機器としては、一般に販売されている機器を利用しました。



図 3-2 Panasonic KV-S4065CWNplustek Opticbook A300

(b) デジタルデータ化における課題と対応策

デジタルデータ化に関する課題と対応策については、以下のとおりです。

① 資料・記録資料・記録加工の必要性

【課題内容と対応策】

資料・記録の中には、加工作業が必要なものがありました。作業量として計画になかったため、作業員を別途アサインする必要がありました。

- ・ 画像方向の修正

画像データで表示向きの修正（縦位置から横位置への修正）が必要なものがありました。

- ・ 音声データの分割

複数の音声ファイルが1つのファイルになっている場合、ファイルを複数に分割する処理が必要でした。

- ・ Web サイトアーカイブ時のトリミング

Web サイトは、画像（JPG 形式）としてキャプチャし保存しましたが、広告部分等アーカイブしない余計な個所をトリミングしました。

(4) 宮城東北大学における実証調査

宮城東北大学プロジェクトでは、次のような視点で資料・記録を収集しました。

➤ 研究者視点での震災関連資料・記録の収集・公開

震災後まもなく、被災地に入り、収集した写真・動画・オーラルヒストリーなど、研究者視点で収集した多くの資料・記録を整備し、公開することにより、防災・減災研究の推進に寄与することを目的としました。

➤ 充実したオーラルヒストリー記録促進

人々の記憶が薄れる前に、被災地域の雇用により、オーラルヒストリーや記録をまとめる作業を組織化し、収集しました。

➤ 東北大学附属図書館の収集資料・記録の公開

東北地方を中心とした図書館グループの震災記録収集活動と連携し、既に電子化された情報を公開可能としました。

以上のような観点で収集した資料・記録において、デジタルデータ化されていない紙媒体（パンフレット、書籍など）や、インタビューの音声などは、大学内の作業場所でデジタルデータ化を実施しました。

また、紙媒体の資料のほか、オーラルヒストリーについてヒアリング記録や写真などのデジタルデータ化を実施しました。

ここでは、紙媒体のデジタルデータ化とオーラルヒストリーの記録手順について紹介します。

(a) デジタルデータ化の手法・プロセス

宮城東北大学では、東北大学附属図書館で収集した刊行物について、デジタルデータ化を実施することになりました。以下に紙媒体におけるデジタルデータ化作業の手順、使用した機器等について紹介します。

【紙媒体のデジタルデータ化事例】

① デジタルデータ化の作業手順

デジタルデータ化にあたっては、作業手順書を作成し、実施しました。以下に作業手順書を示します。

作業手順書 みちのく震録伝 イメージデータ作成フロー		1/1
フロー図	リスク	対応
<p>許諾確認 メタデータ作成</p> <p>許諾可否</p> <p>OK</p> <hr/> <p>イメージデータ作成</p> <p>作業伝票 チェックシート チラシパンフレット 図書・雑誌・古典</p> <p>工程1 底本確認</p> <p>工程2 スキャン</p> <p>工程3 検査</p> <p>NG</p> <p>OK</p> <p>工程4 画像変換</p> <p>工程5 保管</p> <p>作業伝票 チェックシート チラシパンフレット 図書・雑誌・古典</p> <p>コンテナ保管</p> <p>サーバ保管</p> <p>PDF TIFF</p> <p>工程6 記録</p> <p>NG</p> <p>OK</p> <p>チラシパンフレット 図書・雑誌・古典</p> <p>DVD</p>	<p>R1: 伝票と底本名の差異</p> <p>R2: 伝票と底本の数量差異</p> <p>R3: ページの破損</p> <p>R4: スキャン作業による底本破損</p> <p>R5: ブックスキャナの破損</p> <p>R6: ビントズレ</p> <p>R7: スキャン処理漏れ</p> <p>R8: 画像傾き許容範囲超過</p> <p>R9: 画像変換後の数量差異</p> <p>R10: データの紛失</p> <p>R11: 記録エラー</p>	<p>管理者に報告</p> <p>発送工程の作業者確認</p> <p>他の底本と混入した可能性有り</p> <p>イメージデータ作業を全て一旦停止</p> <p>管理者に報告</p> <p>イメージデータ作業を全て一旦停止</p> <p>原因調査し対策立案後に水平展開</p> <p>管理者に報告</p> <p>工程2に戻り再スキャン処理</p> <p>処理の設定内容を再確認、</p> <p>原因不明であれば管理者に報告</p> <p>管理者に報告</p> <p>サーバを再検索し見つからない場合</p> <p>資料IDの底本を再度スキャン</p> <p>原因調査し対策立案後に水平展開</p> <p>管理者に報告</p> <p>書き込みソフト以外のソフトが起動</p> <p>していないか確認</p> <p>書き込み設定内容を確認</p>

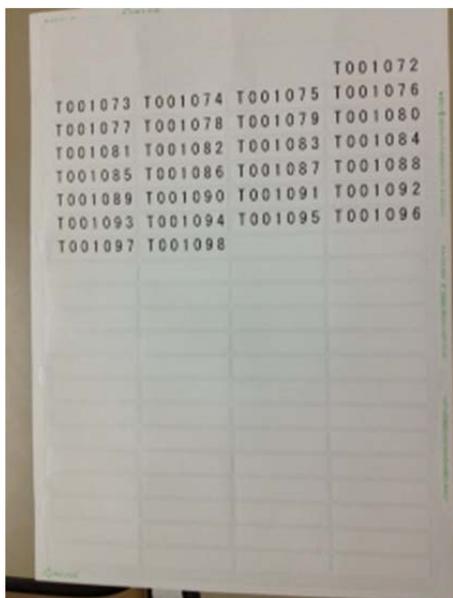
図 3-3 宮城東北大学のデジタルデータ化作業フロー

チラシやパンフレット、古書等の刊行物は、1枚もの、複数枚の冊子もの、図書といった様々な枚数や製本状態のものが存在します。製本された冊子については、1枚ずつデジタルカメラで撮影し、デジタルデータ化を行いました。撮影時には、光の加減により撮影後の濃淡が変わるため、部屋の明かりを一定にする必要があり、窓に暗幕を掛け、光の加減を一定にして実施しました。



環境を一定の光源化とし、撮影前に色調整を実施。

デジタルデータ化した資料・記録は、デジタルデータとの対応を分かりやすくするため、デジタルデータのファイル名をシールに印刷し、クリアファイルに貼ったものに入れて保管を行うこととしました。



② デジタルデータ化における機器等

デジタルデータ化の機器は、以下のものを使用しました。

表 3-16 デジタルデータ化における機器等

機種	台数	備考
Fuji Xerox DocuScan C4250	1	【フラットベットスキャナ】 A3 スキャナ。カラー（フラットベット/シートフィード）、縮小光学系による CCD 固体走査方式。
GRAPHTEC CS500 PRO	2	【大判スキャナ】 A1 スキャナ。最高 4,800dpi の補完解像度と CIS 方式による高画質。本体にモノクロ用の画像処理機能を搭載。
マイクロテック社 BOOK SHOT	2	【ブックスキャナ】 製本された刊行物用のスキャナ。多彩な画像処理機能を搭載し、処理前・処理後と比較しながら作業が行え、専門知識のない方でも直観的に処理が可能。

デジタルデータ化を行う機器としては、一般に販売されている機器を利用することとしました。チラシなどの A4 刊行物等に DocuScan を用い、ポスターなど大判のものは、CS500 を用いました。BOOK SHOT は、図書のような製本されたものに用いました。BOOK SHOT のような製本物専用のスキャナ機器ではない場合には、きれいに読み込むために、製本された図書を裁断する必要があります。作業の際は、BOOK SHOT に限り、光の加減を決める作業を事前に行う必要があります、光源を一定にする工夫が必要でした。



図 3-4 機器等の形状

- ・ポスター等、A3 より大きいサイズのものもあったため、大判スキャナを導入しました。
- ・A3 以下の紙資料は、オートフィーダ機能付きのフラットベットスキャナでデジタルデータ化しました。
- ・書籍等の裁断を行わずにデジタルデータ化可能であること、スキャン後の圧縮率が指定可能であることから、業務用のハイスペックなブックスキャナ（マイクロテック社製 BookShotPro）を使用しました。

これらの機器を以下のように配置し、作業を実施しました。作業風景は、以下のとおりとなります。

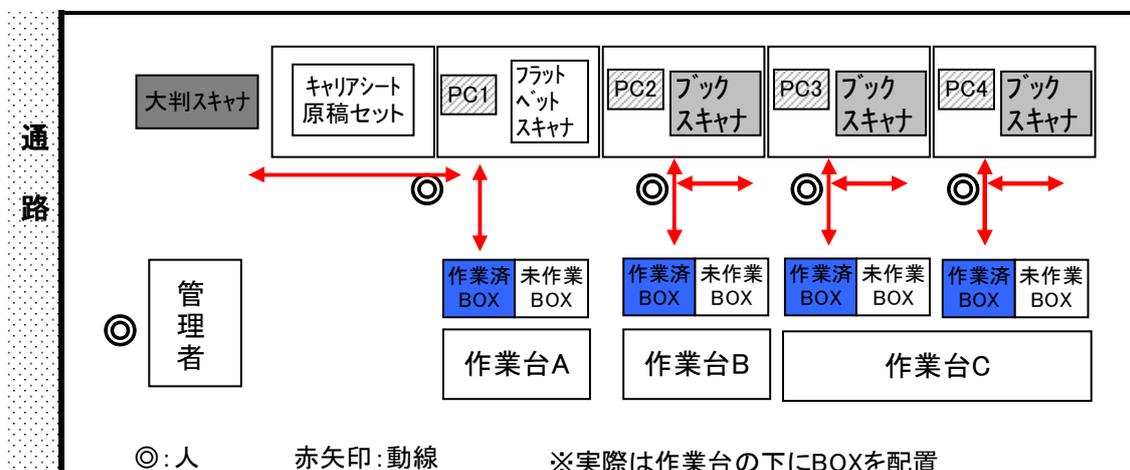


図 3-5 機器配置図



図 3-6 作業風景

宮城東北大学では、紙媒体の他にオーラルヒストリーのデジタルデータ化を実施しました。以下にオーラルヒストリーにおけるデジタルデータ化作業の手順、使用した機器

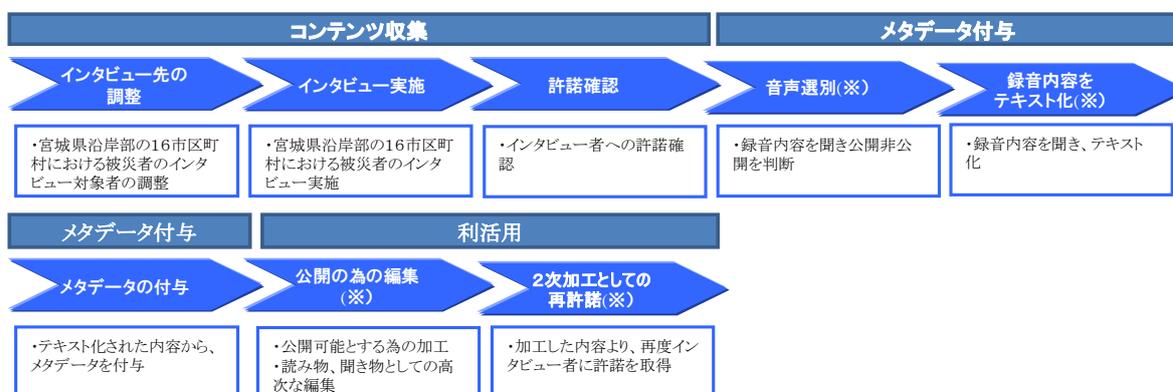
等について紹介します。

【オーラルヒストリーのデジタルデータ化事例】

音声資料・記録収集の具体例として、「みちのくいまをつたえ隊」の活動において、宮城県沿岸部の被災地域在住の方より、オーラルヒストリーをインタビューにより取得しました。インタビューは、被災地在住の方が実施しました。これは、外部の人間がインタビューを行うよりも、現地の人間が行う方が、住民が話をしてくれるという経験からきています。

① デジタルデータ化の手順

デジタルデータ化の手順は、以下に示すとおりとなります。



補足：上記※印の作業は、実証調査として一部作業を行い、後年運用の中でコンテンツ化する予定。

図 3-7 オーラルヒストリー作業手順

インタビューで取得した音声そのものとテキスト化情報については、「非公開」の扱いとしました。その理由は、以下のとおりです。

- ▶ インタビューする側もインタビューされる側も、プロではないため、話が散漫となり、音声の状態ではその場にはいない第三者が理解することができませんでした。音声から内容をまとめ、記事として公開することとしました。
- ▶ インタビューの音声は方言が含まれていたため、内容が理解できない部分がありました。

インタビューで取得した音声及びテキスト情報については、利用者に理解しやすいよう記事として編集し、新たな資料・記録として登録を実施しました。

(参考) みちのく震録伝ホームページより みちのく・いまをつたえ隊活動風景



② デジタルデータ化における機器等

オーラルヒストリーのデジタルデータ化に用いた機器は、以下のとおりとなります。

表 3-17 デジタルデータ化における機器等

分類	製品名	用途
デジタルカメラ	NikonCoolpix AW100	みちのくいまをつたえ隊による被災地の撮影
IC レコーダー	SONY ICD-UX525	インタビューの取得
USB メモリ /HDD	IO-DATA	資料・記録をコピーし、収集用途で使用

機器に関しては、全て、一般に販売されている機器を利用しました。デジタルカメラは、GPS の補足が早く、方位を測定可能な機種を選定しました。IC レコーダーは、ボタンの大きなものを選定しました。収集当時は操作に不慣れなため、困惑した担当者がありました。デジタルカメラについては、GPS 機能が室内や一部の場所で使えないケースが確認できています。

(b) デジタルデータ化における課題と対応策

デジタルデータ化に関する課題と対応策については、以下のとおりとなります。

① デジタルデータ化する際の解像度の決定方法

【課題と対応策】

300dpi の PDF 形式のファイルは、刊行物の枚数が多いと容量が大きくなる可能性があるため、最適な基準を決定する必要があります。本プロジェクトでは試験的に、資料・記録の種類別に複数の解像度でのデジタルデータ化を実施し、それぞれの容量や、解像度について検討を行いました。

変換結果例は、以下のとおりです。

資料ID	資料タイトル	ファイル名	PDF変換設定	サイズ(MB)	圧縮率	サンプル図
R00839	教育いばらき 写真データ 1枚 (チラシ・パンフレット)	R00839_0001.tif	—	40.4	—	
		R00839_0001_HE.pdf	PDF JPEG圧縮 画質:最高	22.5	55.8%	
		R00839_0001_H.pdf	PDF JPEG圧縮 画質:高	4.3	10.6%	
		R00839_0001_M.pdf	PDF JPEG圧縮 画質:中	2.4	5.9%	
		R00839_0001_L.pdf	PDF JPEG圧縮 画質:低	1.7	4.2%	
		R00839_0001_LE.pdf	PDF JPEG圧縮 画質:最低	1.0	2.5%	
R01174	仙台89ERSホームゲーム 無料招待のご案内 文書データ 1枚 (チラシ・パンフレット)	R01174_0001.tif	—	6.4	—	
		R01174_0001_HE.pdf	PDF JPEG圧縮 画質:最高	5.5	85.9%	
		R01174_0001_H.pdf	PDF JPEG圧縮 画質:高	1.4	21.9%	
		R01174_0001_M.pdf	PDF JPEG圧縮 画質:中	1.0	15.6%	
		R01174_0001_L.pdf	PDF JPEG圧縮 画質:低	0.9	14.1%	
		R01174_0001_LE.pdf	PDF JPEG圧縮 画質:最低	0.6	9.4%	

注：「圧縮率」＝「圧縮後のサイズ」÷「圧縮前のサイズ」

写真データ、文章データ等、資料・記録の種類に応じてデータ容量が異なるため、保存に最適な形式は資料・記録の種類毎に決定する必要がありました。宮城東北大学では、容量との兼ね合いから、前述に示した解像度（400dpi）と圧縮率（最低の設定）に決定しましたが、アーカイブにより容量の問題や、データの想定する活用方法が異なることが想定されるため、最適なデジタルデータ化の基準をその都度検討する必要があります。

(5) 宮城河北新報社における実証調査

河北新報社は被災地域の報道機関として、東日本大震災の発生直後から、被災地域での取材活動を継続的に続けており、新聞記事や報道写真をはじめ、膨大な量の震災関連資料・記録を所有しています。また、河北新報社が運営する地域 SNS「ふらっと」に投稿写真サイトを設置し、NPO や学生ボランティアと共に被災地に入り、その様子をブログにまとめるなど、報道とは別の角度からも、一般市民の証言や記録の収集を行いました。

(資料・記録収集方針)

被災地域の新聞社が有する膨大な量の震災関連資料・記録流通の仕組み作り	震災関連の新聞記事の提供（本文は有償）により、震災当時、被災地で何が起こっていたかを新聞記事を通して、把握することができる。また、その新聞記事を補足する写真との関連付けを表現することにより、将来の防災・減災研究の促進に寄与する。
既存システム（商用）と震災関連デジタルアーカイブ（無償）との	マスメディアが所有する既存の本番商用データベースと、本実証事業で構築する震災関連デジタルアーカイブ

有機的な結合と境界線の確立	ブとの連携を、新聞記事（見出し、タグ情報など）の自動配信の仕組みにより有機的に結合する。また、有償資料・記録と無償資料・記録の境界線を明確にする。
報道機関の強みと、被災地域の地の利を生かした、広い震災関連資料・記録収集活動の推進	報道機関としての機動力と、地域コミュニティとのリレーション、被災地域の地の利を最大限に活用し、地域で活躍するボランティア団体などからの資料・記録収集や、一般市民からの投稿促進のための施策により、幅広い震災関連資料・記録の収集が可能。

収集された資料・記録は、ほとんどがデジタルデータ化されていたため、デジタルデータ化作業は発生しませんでした。

ただし、河北新報社では、1991年以前の新聞記事をマイクロフィルムで保存しており、これをアーカイブコンテンツとすることとしました。

ここでは、そのデジタルデータ化の状況について、紹介します。

(a) デジタルデータ化の手法・プロセス

河北新報社では、1991年以前の新聞記事についてマイクロフィルムをデジタルデータ化し、記事のテキスト化を実施しました。その際の手順を以下に紹介します。

① デジタルデータ化の手順

デジタルデータ化の全体のプロセスは以下のとおりとなります。



図 3-8 新聞記事のデジタルデータ化作業手順

新聞記事のデジタルデータ化は、マイクロフィルムより記事単位に切り出し、PDF形式にする作業と記事本文をテキスト化する作業となります。

当初、テキスト化作業は、OCR機能を利用し、イメージデータからテキスト化を実施する予定でしたが、新聞記事は連続した文字列として認識されないことと、古い紙面では、旧字体が多く存在し、認識率が低かったことから、紙面を見ながら文字をテキスト入力することとしました。

② デジタルデータ化における機器

システムは通常の業務で新聞記事の紙面を構成する時に使用している組版システムを使用しました。マイクロフィルムをスキャンし、イメージデータを作成する作業は、外部の会社に委託しました。

(b) デジタルデータ化における課題と対応策

河北新報社におけるデジタルデータ化作業に関する課題と対応策は、以下のとおりとなります。

① OCR の認識率について

【課題内容と対応策】

OCR でテキスト化した結果、連続した文字列でテキスト化できないと認識率が低くなるのが分かりました。また、新聞紙面では連続した文字列とすることが困難である上、古い紙面は旧字体がかなり存在し、誤認識がかなり発生しました。

そこで、OCR ソフトで文字をテキスト化することは断念し、新聞紙面を見ながら手入力することとしました。

② マイクロフィルムの劣化による対応

【課題内容と対応策】

マイクロフィルムの状態が良くないものがあり、記事のイメージ化の際、シミ等にマスキングを行い、体裁を整える作業を実施しました。

(6) 福島県における実証調査

福島プロジェクトでは、主に自治体などを対象に、市民向けの情報や報道発表資料を中心に資料・記録収集を行うこととしました。

資料・記録収集は、原則として公開されたデジタル情報を対象として行うこととしたため、デジタルデータ化が必要となったものは、データの残っていない広報誌や文書類の一部のみでした。ここでは、いわき市より受領した行政文書をデジタルデータ化した際の事例と南相馬市が運営するケーブルテレビ局から受領した映像データのデジタルデータ化事例について紹介します。

(a) デジタルデータ化の手法・プロセス

【いわき市プレリリース資料のデジタルデータ化事例】

① デジタルデータ化における機器等

いわき市プレスリリース資料は 246 点あり、デジタルデータ化に使用した機器等は、以下のとおりとなります。

表 3-18 デジタルデータ化における機器等

分類	メーカー名	型番 等	備考
画像スキャニング (単票)	キャノン	CanScanLiDE200	一般ユーザ向け
画像スキャニング (冊子等)	plustek	Opticbook A300	フラットベッド型 A3 サイズまで対応 標準附属ソフト利用

【ケーブルテレビ局からの映像資料のデジタルデータ化事例】

① デジタルデータ化の手順とデジタルデータ化形式

ここでは、ケーブルテレビ局等の業務用映像資料、記録映像資料をデジタルアーカイブで公開するためのフォーマット変換、ファイル分割の方法について紹介します。

- ・ 南相馬市が運営するインターネット放送局南相馬チャンネルの番組データについては、フルハイビジョン番組で1番組あたりのサイズが最も大きいものでは30分番組で3.5GB程度のものもあり、デジタルアーカイブで閲覧する際のダウンロード時間等を考慮するとサイズを圧縮する必要性がありました。
- ・ サイズの圧縮方法としては、ファイルの分割によるものと、画質・音質を下げることによるものの大きく2通りが考えられます。動画のWeb配信に関しては、Youtubeにおける動画のアップロードに関するガイドライン²⁶を参考にして動画のサイズ圧縮方針を決定しました。具体的には、動画の再生時間については1ファイル15分を上限としました。また、動画の変換方式に関しては、以下のような設定を採用しました。

表 3-19 動画の変換方式

設定	設定内容
コンテナ	4MP4
音声コーデック	AAC-LC
動画コーデック	H.264
解像度	1280×720 (オリジナル)
ビットレート	5,000kbps (標準画質)

26

(<http://support.google.com/youtube/bin/answer.py?hl=ja&answer=71673&topic=2888648&ctx=topic>)

- 南相馬チャンネルの動画については、ファイル形式が **MOV** 形式となっており、エンコードにより **MP4** 形式に変換することによってファイルサイズが約 4 分の 1 に圧縮されるため、15 分の動画であればファイルサイズは約 **400MB** 程度となることが判明しました。
- ラジオの音声データについては、オリジナルの音声ファイルは **WAV** でしたが、この形式だと 1 時間の音声ファイルが **600MB** 程度とファイルサイズが大きくなってしまったため、音声圧縮形式のひとつである **MP3** に変換することで約 10 分の 1 に圧縮してアーカイブに格納することとしました。これは、実際に南相馬ひばり **FM** で音声ファイルの保存用に実施している手法であり、これを福島プロジェクトでは参考としました。