

ABCC (原爆傷害調査委員会) 文書を素材にした史料画像デジタル・アーカイビング・モデルの作成に関する研究

大学院人文科学研究科アーカイブズ学専攻教授 安藤 正人
 学習院大学計算機センター、大学院人文科学研究科アーカイブズ学専攻教授 入澤 寿美
 学習院大学文学部史学科教授 千葉 功
 学習院大学大学院人文科学研究科アーカイブズ学専攻助教 青木 祐一

1. 研究の目的

本研究は、広島・長崎原爆被爆者の放射線被害に関する、医学データを含む調査記録、研究記録、行政記録等の資料を、日米両国を中心に広く収集し、これをデジタル・アーカイブズとして公開利用に供することを最終目的とした大きな研究の一環である。具体的には、平成21年～24年度科研費補助金基盤研究(A)「旧日本植民地・占領地関係資料ならびに原爆関係資料のアーカイブズ学的研究」(研究代表者・安藤正人)によって撮影・収集した米国科学アカデミー所蔵「原爆傷害調査委員会(Atomic Bomb Casualty Commission, ABCC)文書」のデジタル画像約14万コマを素材として、アーカイブズ学的な史料構造分析にもとづく史料画像デジタル・アーカイビング・モデルを作成することが目的である。

広島・長崎の原爆放射線被害については、原爆投下直後より現在に至るまで、日米両国を中心に官民間問わず多くの調査や研究が行われてきた。広島と長崎に置かれた日米共同の調査研究機関「原爆傷害調査委員会 ABCC」は、その中心である。研究成果は、これまで放射線医療や原子力産業など多くの分野で活用されてきた。しかし、2011年3月の福島第一原子力発電所事故を機に、改めて放射線の人体への影響に関する過去のデータに関心が集まった結果、広島・長崎被爆者の放射線被害調査データや医療データなどの記録が、必ずしも十分に研究資源化されておらず、今回のような事態に際して迅速かつ効果的に活用できるシステムが整っていないことが浮き彫りになった。

このような状況を改善するため、私たちは以下のことが必要と考えている。すなわち、第一に、日米両国ならびにアジア近隣諸国に残存する関係資料の徹底的な調査を実施すると、第二に、明らかになった資料所在情報、ならびに主要資料の画像データを含む内容情報を、アーカイブズ学的方法にもとづいて適切に蓄積し、これを研究資源として世界に発信する「原爆放射線被害デジタル・アーカイブズ(仮称)」の構築である。本研究は、上記にいう第二の研究の一環であり、史料画像のデータベース構築とその公開システムの開発に関わる問題に特化した基礎研究を行おうとするものである。さいわい代表研究者の手もとには、米国科学アカデミーから収集した ABCC 文書約 14 万コマという最適の素材がある。これは ABCC の米国側本部資料にあたるもので、内容が豊富か

つ高質であり、原爆放射線被害に関するもっとも重要な史料群のひとつである。本研究を通じて、学術研究資料の整備という点にとどまらず、広く現代の社会的ニーズにも応えていきたいと考えた。

なお、「原爆放射線被害デジタル・アーカイブズ（仮称）」の構築という将来的な目標については、科学研究費補助金基盤研究（A）「国際コンソーシアムによる『原爆放射線被害デジタルアーカイブズ』の構築に関する研究」（平成25年度～28年度、研究代表者・安藤正人）を実施することになった。よって、本研究はそれと密接に関連させながら進めることとなった。

2. 研究の概要と成果

1年間という限られた期間で有効な研究成果をあげるため、本研究では、第一に研究素材を米国科学アカデミーから入手したABCC資料のデジタルデータに絞ること、第二に二つの研究チームを設けて、それぞれ課題を限定した専門的調査・研究を実施することを基本方針とした。二つの研究チームは「メタデータ研究チーム」と「個人情報研究チーム」である。各研究チームには、研究メンバーのほかに、外部の専門研究者ならびに学習院大学大学院人文科学研究科アーカイブズ学専攻の大学院生など多数の協力者が参加してくれた。主な協力者の氏名は次の通り。和田華子、芹澤良子、広瀬真紀、平野泉、橋本陽、元ナミ（なお上記科研費基盤研究(A)に関わる研究分担者、連携研究者、研究協力者の皆さんにもお世話になったが、全員のお名前をあげられないので省略することをお許しいただきたい）。

各研究チームはそれぞれ、日常的な調査活動と数回の個別の研究会を行った。「メタデータ研究チーム」は、科研費による米国医療センター・アーカイブズとの共同研究会にも参加した。また、全体研究会を2013年5月19日（学習院大学）、9月15日（学習院大学）、11月24日、25日（広島県立大学、広島大学放射線医科学研究所ほか）、2014年2月22日（学習院大学）の4回開催した。いずれも科研費研究プロジェクトとの連携の下に行われたが、11月24日、25日の広島研究会は主として本研究プロジェクトの事業として実施したので、以下この研究会における研究成果報告の内容を中心に記述する。

（1）メタデータ研究チーム

米国科学アカデミーより収集したABCC資料のカラー画像約14万枚について、アーカイブズ学的記述にもとづく目録データベース化と画像データの公開を適切かつ効果的に行う方法を研究するため、「メタデータ研究チーム」は、米国におけるアーカイブズ情報システムの調査と研究、ABCC資料の画像分析、画像情報のメタデータ要素とその標準化に関する研究を担当した。主なメンバーは、入澤寿美、平野泉、橋本陽、元ナミの諸氏である

まず米国（北米）におけるアーカイブズ情報システムについて、デジタル化資料搭載のウェブサ

イトを調査し、埋込型 (Embedded Model: 目録から辿る形式)、分離型 (Segregated Model: 1 件レベル検索方式)、並立型 (Parallel Model: 埋込型と分離型の複合形式) の 3 種があることを紹介し、埋込型は目録中心の視点から一面的な検索を行うもので一件あたりの検索性に問題があること、分離型は多面的検索を可能にするが資料群のコンテキストが見えないこと、並立型は前二者の欠点を補完しようとしているが結果的にユーザーが片方しか使わないという問題がある、などの特徴を紹介した。またコストカットの面では埋込型が有利であり、分離型はメタデータ付与に大きな労力が必要な点なども指摘した。

デジタル・アーカイビング・モデルの構築については、上記のような北米における諸類型をも参考にしながら、国際標準 (EAD: Encoded Archival Description が基本) にもとづいた ABCC 資料の目録記述モデルと画像情報のメタデータ記述モデルの作成に向けた更なる研究が必要になる。これについては、2013 年度後半にテキサス医療センター・アーカイブズとの共同研究が進み、その結果、open source archival description software である ICA A to M の活用など一定の方向性が決定された。そして、2014 年度にかけて ABCC 資料の一部を素材にしたパイロット・プロジェクトを実施することになり、既にその作業に着手したところである。

(2) 個人情報研究チーム

ABCC 資料のデジタル画像を将来的に公開するにあたって、デジタル・アーカイブズの構築に関わる理論的・技術的問題とともに極めて重要になるのが、画像に含まれる個人のプライバシーに関わる情報をいかに保護するか、という問題である。ABCC 資料には、個人を特定できるもの、できないものを含め、被爆者に関する多くの医学データや写真が含まれており、それらをどのように取り扱うかは、容易に解答が見出せる問題ではない。日米におけるプライバシーについての考え方や慣習の違いも存在する。そこで、本研究チームは、アーカイブズ資料に含まれる個人のプライバシー関連情報の保護に関する理論と方法を研究し、最終的に「ABCC 関係資料におけるプライバシー問題への対応に関するガイドライン (仮称)」を作成することを目的として、第一に国立国会図書館等において内外の基礎的文献を調査・分析するとともに、第二に個人情報チェックマニュアルを試作し、ABCC 資料の一部を用いて実験的に個人情報チェック作業を行った。以下では上記第二の点について、研究成果を略述する。

試験的チェック作業を行った資料画像は約 500 点で、文書、写真、地図などが含まれる。資料の種別は以下のように大きく H と A の 2 種類に分け、あらかじめ作成した個人情報チェックリストに照らし合わせながら、該当する個人情報の有無によって H1 ~ H3、A1 ~ A3 の記号を与えた。

H：被爆者（死亡者も含む）及びその家族（遺族）に関連する資料

H1：被爆者及びその家族（遺族）に関連する個人情報を含まない。または、被爆者及びその家族（遺族）に関連する個人情報を含むが、個人を特定する情報（氏名・住所等）を含まない。

H2：被爆者及びその家族（遺族）に関連する個人情報を含み、かつ個人を特定する情報を含むが、マスキング等の匿名化が可能。

H3：被爆者及びその家族（遺族）に関連する個人情報を含み、かつマスキング等の匿名化が不可能。

A：ABCC 及び NAS の職員に関する資料

A1：ABCC 及び NAS の職員に関連する個人情報を含まない。または、ABCC 及び NAS の職員に関連する個人情報を含むが、個人を特定する情報（氏名・住所等）を含まない。

A2：ABCC 及び NAS の職員に関連する個人情報を含み、かつ個人を特定する情報を含むが、マスキング等の匿名化が可能。

A3：ABCC 及び NAS の職員に関連する個人情報を含み、かつマスキング等の匿名化が不可能。

試験的チェック作業の結果、個人情報を正確に抽出するためには、資料作成主体である ABCC や NAS（米国科学アカデミー）についての知識はもちろん、資料のほとんどが英文の専門資料であることから、英文読解能力と医学・病理学等の専門用語についての知識もある程度必要であることが改めて明らかになった。また、ABCC 職員や病理学者等の氏名・配属・時給など、非公開の対象にすべき個人情報といえるかどうか検討を要する具体的問題が多数あることも明らかになった。

まとめ

本研究は、「原爆放射線被害デジタル・アーカイブズ」の構築を最終目標にした大きな研究の一環であり、その中で、米国科学アカデミーが所蔵する ABCC（原爆傷害調査委員会）資料のデジタルデータを素材に資料画像デジタル・アーカイビング・モデルの作成を試みようとする研究プロジェクトであった。1 年間という時間的制約のため資料画像のアーカイビング・モデルを完成させるには至らなかったが、国際標準にもとづいた目録記述と画像情報のメタデータ記述のモデリングについて、具体的な方向性を絞るところまで研究を進めることができ、平成 25 年度末からテキサス医療センター・アーカイブズとの ABCC 資料を素材とした共同パイロット・プロジェクトを発足させることになった。このパイロット・プロジェクトは、上記科研費を活用して平成 26 年 10 月までに終了させ、次のステップに進む計画である。また「原爆放射線被害デジタル・アーカイブズ」の公開にあたって極めて重要となる個人情報、プライバシー保護の問題に関しても、ABCC 資料を素材にした試験的研究の結果、ガイドラインの作成に向けて一定の成果を得ることができた。以上のような本研究の研究成果は、今後、上記科研費研究を中心に引き継がれる予定である。