

センターの概要

研究資源センター（RRC）は、放影研のインフラの中核的な構成要素となる。放影研の戦略計画を推進するためには、RRCを上手く実現し運営することが不可欠である。

RRCには以下の3つの使命がある：

1. 放影研の研究資源を保護し、索引付けし、統合する。これには、データ、バイオサンプル目録、紙の記録、人工物、原稿、データセット、プログラミングスクリプト、その他歴史的に重要な論説などが含まれる。研究データへのアクセスは、対象者のプライバシーを保護するための明確なアクセス基準を備えたウェブポータルを介して行われる。
2. 全てのデータとバイオサンプルの目録とを統合することにより、放影研の研究遂行能力を高める。データ可視化、データ整理および分析のためのツールにより、アクセス手順を簡略化、標準化し、研究を促進する。
3. 放影研の資源を活用・共有し、所内の研究および契約や外部資金を含め共同研究プロジェクトを促進するための管理体制を提供する。

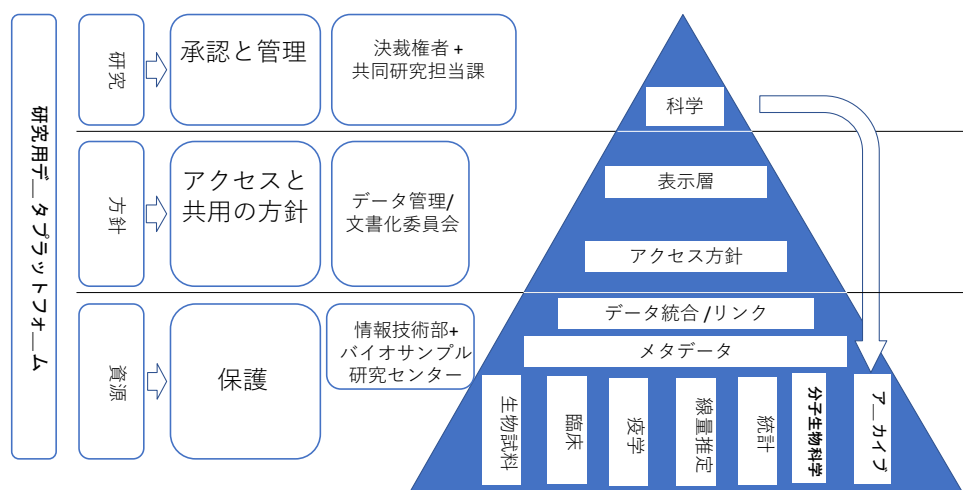


図1. 放影研の研究資源センターの機能的構造

図1を見ると、ピラミッドの最下層は保護層である。保護層には、デジタルデータの保存、統合、可視化のための枠組みが含まれる。デジタルデータ以外にも、この層にはデータ管理システムや、放影研資料のデジタルスキャンした記録や歴史的に重要な人工物の目録を全てカタログ化し索引付けするためのコンテンツ管理システム（CMS）が含まれる。データのコピーやスキャン画像は一元的に保管される（使命1）。全ての資料には各資料を説明する情報（メタデータ）が付くことにより、見つけられ、関連資料と適切に紐付けられ、検索可能になる。第2層は、「方針とアクセス」の層である。所内規則と最新の認証方法を用いて、権限を与えられた資料のみが閲覧可能となる。この層には、所内外の研究者が放影研のデータカタログを閲覧できるようなプレゼンテーションツールが含まれる。権限を付与されたユーザーは、これらのデータに基づく研究計画書や論文原稿を閲覧し、新たな研究を計画するためにデータを可視化し検索することができる（使命2）。最上層は研究層である。ここでは、承認された研究を実施できる。共同研究を行う場合は、「研究支援室（ORS）」がデータを提供し、バイオサンプルの提供依頼を処理する

ための契約を履行するとともに手続きを支援し、研究を円滑に進める（使命3）。研究終了時には、当該研究に関連する全ての資料（データ、分析結果、解析スクリプト、論文原稿など）を適切なメタデータを付与しリンクを立て、アーカイブに保存し、将来の研究者がそれらの成果物を再利用できるようにする。

RRCを設立・管理するための人員や機能別ユニットは、研究資源センター運営委員会が暫定的に設定した。この臨時委員会は2021年11月に発足、12月に第1回目の会合が開かれた。Eric Grant主席研究員が委員長および田邊主席研究員が共同委員長として当該委員会を主導し、部長全員が委員を務める。この委員会は、2019年12月に設置された「研究資源センター設置準備委員会」という以前の臨時委員会に取って代わるものである。

3つの主要な開発ニーズ

図1に示すような機能を持つRRCを実現するためには、大きく分けて3つの領域の開発を完了する必要がある。この3つの領域とは、「データフレームワーク」、「コンテンツ管理システム」、「研究支援室（ORS）」である。この機能と開発領域については、放影研の「技術チーム」が2017年以降隔週で会議を開催し綿密に検討してきた。このチームには、ITDの幹部職員とGrant主席研究員が含まれる。

データ管理フレームワーク：「データ管理フレームワーク」は放影研が保有する全ての電子データに関して基盤となるべき計算機資源環境である。データの保存、アノテーション、統合が可能でなければならない。FAIRの原則に従って新しいデータを容易にシステムに追加できるよう、フレキシブルなものが理想的である。（コンピューターに詳しい管理担当者にこの作業を強いるのではなく）データを作成した研究員がアノテーションできれば理想的である。このシステムは日本語と英語の両方に対応し、ローカルサーバーまたはクラウド環境にインストールすることが可能で、データが研究所を「離れる」ことがないよう研究者がリモートで解析を実施できるならば、理想的である。また、認証された人だけがデータにアクセスできるよう、階層的で安全性の高い認証システムが必要である。またデータを可視化するとともにデータ探索のために検索できる、「データ可視化アプリケーション」が基礎データと相互作用できるようにする。

このようなシステムに関するパイロットプロジェクト（シカゴ大学が開発したGen3データコモンズ）を2020年から2021年に1年以上かけて実施した。残念ながら、本章で後述する様々な理由により、このパイロットプロジェクトは成功しなかった。放影研は現在、適切なデータ管理フレームワークを模索している。これは、RRCが直面している現在の最大の課題の一つである。

現在、データ可視化ソフトを使ったパイロットプロジェクトが進行中で、SACに対してデモンストラーションを行う予定である。

コンテンツ管理システム：コンテンツ管理システム（CMS）は、放影研内の全ての紙ベース、またはテキストベースの資産を保存するシステムである。放影研の認可アクセス方針を厳守しながら、放影研資料へのアクセスを可能にするフレキシブルで頑健なセキュリティシステムが必要である。より一般的な資料（出版物リストなど）については、現在、放影研には様々なリスト（研究計画書、業績報告書など）を閲覧するための多数のデータベースが存在するが、それらはしばしば独立したデータベースであり（すなわち、複数のサイトを検索して調べる必要がある）、放影研内のどこにも全文検索機能は存在しない。これらの問題（一元的なアクセスや全文検索が不可能）は、放影研の現在の計算機資源環境における大きな欠点である。放影研内の紙ベースの記録、放影研の全てのPDFファイル、検査記録のノート、質問票などをスキャンして保存できるCMSが、放影研には必要である。このシステムは日英に対応し、直感的な操作が可能であり、保存と再利用のために個

人がローカルで保有された資産をアップロードできるものでなくてはならない。

パイロットプロジェクトを策定するため、RRC技術チームはCMSを特定し、当該CMSを専門とする外部業者を起用した。SACに対しこのシステムのデモンストレーションを行う予定である。

研究支援室：放影研の研究活動は、管理面の障壁によって不必要に遅れが生じている。全ての業務が紙ベースで行われ、他の承認書と重複または矛盾する承認書も多い。さらに、外部の研究者が放影研の保存資料を閲覧したり、共同研究プロジェクトを検討するための打合せを始めるための「プレゼンス」情報が、放影研の外部ホームページにはない。研究支援室（ORS）は、これらの問題を一つ一つ解決することを目指す。放影研の申請書を徹底的に検討し、重複を避けるために書式を作り直すことを想定している。さらに、申請等の関連プロセス全体をオンラインシステムに移行すべきである。このシステムを、カスタマイズされたワークフローを作成できる企業資源計画（ERP）パッケージを用いて構築する。日本語および英語に堪能な新たな人員を雇用もしくは登用する必要がある。

人件費、請負業者の支援、ソフトウェア調達を含む暫定予算が策定された。

2021年度業績

2021年の研究資源センターの主要な業績は、管理体制の再整備であった。

当初、RRCに関する計画は、2019年12月に設立され、丹羽理事長およびUllrich副理事長が共同委員長を務めた「研究資源センター設置準備委員会」が主導していた。この臨時委員会は、全部門に情報を提供し、支持を取り付けるために設立された。この臨時委員会の中で、「ガイドライン起草」小委員会は、報告系統、リーダーシップ、予算、人員、RRCの内部構造など、放影研におけるRRCの位置づけを決定する役割を担った。第5回会議は2021年に開催され、その時までには目録作成やスキミングの準備に関わるいくつかのプロジェクトを完了した。Eric Grant主席研究員を委員長、田邊主席研究員を副委員長とする新たな臨時委員会「研究資源センター運営委員会」（RRCOC）の設置に伴い、2021年11月にこの臨時委員会は廃止された。RRCOCの第1回会議は2021年12月に開催された。

RRCは当初、物理的にも独立し独自の人員を有する放影研内の独立した「センター」となることを計画していた。現在でも長期的にはそうなる計画である。しかし設立を迅速に行うため、まず放影研の既存の部署にいくつかの新しい課を設立し、その後それらの課をRRCとしてまとめる「2段階」プロセスが採用された。「概要」で述べたように、RRCが機能するために必要な主な開発領域が3つ存在し、それは、1) データ統合、データのタグ付け、データ可視化、データセットの整理のためのデータフレームワークの設置および運用、2) 放影研内の全ての文書を一元管理し、（日英両方での）全文検索を可能にするコンテンツ管理システム、3) 全ての申請／承認手続きを合理化し、これらの手続きをオンラインシステムに移行させるとともに、RRCのオンラインプレゼンスを確立することを任務とする研究支援室、である。これらの課題を達成するために、RRCOCは以下の3つの課を新たに設立する（または既存の課を当該目的に充てる）ことを承認した。

- 研究資源課（RRS）：この新しい課は、情報技術部内に設置される技術面を担当する課で、小野部長が主導する予定である。この課は、データのタグ付け、データの視覚化、研究用の新しいデータセットの整理を含む放影研のデータを一元管理するためのフレキシブルなシステムを提供する「データ管理フレームワーク」の設置を監督することを任務とする。RRSでは、まずRRCのプログラミング作業を指揮する上級プログラマーを契約スタッフとして採用する。2023年度には、新しいプログラマーをRRSに採用する予定である。さらに、ITDのシステム技術課の職員2名にRRSのプログラマーを兼任させ、データ統合と開発作業を支援することにより、

RRSを更に強化する予定である。

- 図書資料課：これは放影研に既存の課だが、新しいコンテンツ管理システム（CMS）に論文、研究計画書、業績報告書などを識別し、アップロードし、カタログ化する役割を追加する。また、外部業者とも協力して、放影研のニーズに合わせてCMSをカスタマイズする。この課は小野部長に直属する。
- 研究支援室（ORS）：ORSを放影研事務局内に新設する。事務局は規則や方針の実施に精通している。（日本語および英語に堪能な）新しい職員1名を採用もしくは登用し、事務局長直属とする。この職員は、事務管理の負担を軽減するために放影研の研究手順の全面的な見直しを指揮する。この目的は、所内の研究および共同研究の両方を促進するために、申請手続きと承認手続きの合理化、すべての手続きのオンラインシステムへの移行、所内外の研究者へのWebプレゼンスの確立などにより達成される。また、事務局内に設置されるため、研究助成金の申請やその他の政府の求める提出要件を支援する上で会計課と密接に連携することが可能になる。

RRCOCがRRCの今後の体制や人員に関する要件を設定したことに加え、いくつかのプロジェクトが進行中である。

- RRC設立に向けた5か年スケジュール：Grant 主席研究員および小野部長が協力し、RRCにおける技術開発のスケジュールを作成した。
このスケジュールに含まれる主要な項目は、技術的發展に必要な放影研のインフラに関するいくつかの整備（多要素認証を含む放影研の全リソースに対するシングルサインオンシステムなど）、現在放影研内で手作業と紙ベースで行われている多くのプロセスの自動化と合理化を可能にする企業資源計画（ERP）システムの導入である。その他の主な項目としては、上記の業務を扱う新しい課の設立スケジュールも含まれる。RRCは2024年度末までに外部向けのオンラインプレゼンスを確立し、2025年度末までにフル稼働する予定である。
- スキャニングセンターに関するパイロットプロジェクトの始動：小野部長の指導の下、ITDは放影研の多目的コピー機の多くを、放影研の「スキャニングセンター」用に再構築した。スキャニングセンターは、個々の研究者や部署が個人的なノートや論文のデジタルフォーマットへの移行に着手できる手段である。スキャナーにログインすると、全てのスキャンデータは光学式文字認識機能を通して送信され、新しく作成されたPDFファイルは作成者の名前で一元的に管理される。その後、これらのファイルを再編成し保管することができる。最終的には、これらのファイルは自動的に放影研のCMSシステムに移行しタグ付けされ、適切な（そして変更可能な）アクセスプロファイル（誰が当該文書にアクセスできるか）が割り当てられる。
- コンテンツ管理システムに関する予備調査：放影研はコンテンツ管理のためのオープンソースソフトウェアを特定し、そのソフトウェアを専門とする日本で有数のプログラミング企業を探し出した。そして、放影研のコンテンツ管理のためのパイロットシステムを開発するため当該企業と契約を締結した。作業は進んでいるが、放影研の設計仕様は相当難易度が高いものであることが判明した。その仕様とは、放影研のシングルサインオンシステムと互換性のある階層的なアクセスシステム、日本語および英語での完全な対応が可能であること、さまざまな資料の間の関係性を示す複合的なリンク、フレキシブルであること、カスタマイズ可能であることなどである。このシステムが放影研のニーズを満たしてくれると期待するが、まだ「予備調査の段階である」と考えるべきで、成功が保証されているわけではない。
- データ可視化に関する予備調査：放影研は、データ可視化ソフトウェア「Tableau」のライセ

ンスを限定数購入した。これはデータ可視化分野の先進的なソフトウェアであり、放影研のデータを探索するための「フロントエンド」として使用することができる。どんな複雑なソフトウェアでもそうだが、習得にはかなりの時間がかかる。このソフトウェアは有望と思われるが、成功するには、放影研で使用するためのパッケージの習得および開発に専念する人員が必要である。注：SAC会議の際にデモンストレーションを予定している。

- 科学技術振興機構（JST）の戦略的創造研究推進事業補助金を申請：広島大学原爆放射線医科学研究所（原医研）と共同で戦略的創造研究推進事業（RISTEX）を申請した。この外部資金の申請は最初の資金調達の難関を突破し、第二段階の面接に進んだものの、最終的な交付には至らなかった。放影研と原医研の共同申請は、一般市民と社会科学者が関心を寄せる歴史的資料を共同で保有し、一般の人々がアクセスできるようにするためのウェブベースの一般向けアーカイブシステムを構築するというものであった。このシステムは、単一のインターフェイスで検索と提示を行う連合データベース（各研究所がどの資料を提供するかを独自に管理できる）を有するよう設計された。このシステムの設計から得られた知見を、RRCの創設・設計に活用することを意図していた。
- 劣化したマイクロフィルムのデジタル化：原簿管理課および臨床研究部に保管されている数十万枚のマイクロフィルムは、数十年にわたる不適切な保管状態により劣化が進んでいた。これらのフィルムを特定し、外部業者が全てデジタル化した。復元率はおおむね良好で、自動化装置でデジタル化できないほど劣化したリールは5本のみであった。これらのリールについては専門のリペア業者に依頼して可能な限り復元を行った。その結果、物理的に復元できなかったフィルムは数十コマのみであった。
- "Gen3" データコモンズパイロットプロジェクトが終了：シカゴ大学の Robert Grossman博士の来所後、放影研はシカゴ大学が開発し米国国立がん研究所が使用するGen3データフレームワークを用いたパイロットプロジェクトを試みることにした。Gen3は、大規模データの保存と解析のためのオープンソースのソフトウェアプラットフォームである。このパイロットプロジェクトが検討された時点では、放影研は専門知識も人材も持ち合わせていなかった。Grant主席研究員は人脈を用い、テキサス大学メディカルセンターで疫学の博士号を取得中のコンピューター科学者を放影研でのインターンシップに採用した。当初の提案では、当該プログラマーは春にシカゴ大学で数週間の研修を受けた後、夏にはGen3の専任としてインストール作業を行うため広島研究所に来所することになっていた。しかし、新型コロナウイルス感染拡大の影響で、全ての来所は中止せざるを得なかった。彼の名誉のために言っておくと、そのコンピューター科学者は、夏の間および2020年12月にかけて、毎週Grant主席研究員とテレビ会議を行っていた。残念ながら、このパイロットプロジェクトは終了した。これは、個人の過失ではなく、悪い状況が重なってしまったためである。複合的な要因としては、Gen3システムは始まったばかりで文書化が不完全であったこと、Gen3の技術員が日々の業務に追われて放影研のGen3のインストールに十分な時間を割けなかったこと、大学院生がこのプロジェクトを遠隔で行おうとしたことなどが挙げられる。RRC技術チームは、大規模で専門的なシステムの導入と管理には技術チームが必要である

と理解した（最小限のスタッフによる「わずかな」努力では済まない）。また、データフレームワークの開発状況も慎重に評価する必要がある。成熟した状態であればあるほど、導入は容易である。新しく、急速に進化するシステムには、開発チームと密接に連携する技術者が必要である。

技術チームの隔週会議：Grant主席研究員、ITD主導陣、システム技術課および図書資料課の職員は、RRCの様々な業務について詳細に調査し実施計画を立てるため、2019年から隔週で会議を開催している。会議で協議される内容は、高度に概念的な議論から具体的な機能まで多岐に渡っている。技術的なニーズや放影研が望む機能を現在実践しているグループやソフトウェアプラットフォームを特定するため、かなりの時間を費やし調査した。このチームは技術的な問題に焦点を当てている。これらの会議は、正式な委員会の範囲を超えて継続する。2020年に実現しなかったが、RRCの設立に関する取り組みは2017年に始まった。2020年以前に行ったその他の活動、成果を簡単に以下に述べる。

1. 世界中のデータやアーカイブの専門家を訪問。ITDの職員は、データ統合やアーカイブのための既存の専門知識をもともと持っていないため、選択肢や開発戦略について把握するため、大規模プロジェクトに携わっている様々な機関の専門家と連絡を取る必要がある。以下の専門家/機関と対面の会議を開いた：

- Dave Thompson / ウェルカム図書館 デジタルキュレーター
- Stephen Chanock / 米国国立がん研究所 がん疫学・遺伝学部長
- Anthony Philippakis / ブロード研究所 最高データ責任者
- James Cuff / ハーバード大学 研究コンピューティング部門担当副学部長
- Brian White / ハーバード大学 研究コンピューティングITオペレーションマネージャー運用責任者
- John Quackenbush / ダナ・ファーバー癌研究所 生物統計学・計算生物学担当部
- Meir Stampfer / ハーバード公衆衛生大学院 看護師健康調査研究責任者(PI)
- James Lacey / カリフォルニア教師研究 健康分析論分析学責任者
- Tom Murphy / ミシガン大学 政治・社会調査のための大学協会 (ICPSR) コンピューターおよびネットワークサービス部長
- Robert Grossman / シカゴ大学 生物科学部 研究情報科学最高責任者
- 有賀 暢迪 / 国立科学博物館 理工学研究部
- 杉本 重雄 / 筑波大学 図書館情報メディア研究科

2. James Cuff博士（ハーバード大学）の来所、2017年5月

放影研の研究上のニーズと現在のインフラを評価するために、ハーバード大学研究コンピューティング部門担当副学部長であるJames Cuff博士が3日間の日程で来所した。Cuff博士

はITDならびに放影研の主導陣と協力し、問題を調査し論理的な手順段階に分類解するために「白書」の概要を作成した。

3. Robert Grossman博士（シカゴ大学）の来所、2019年1月

シカゴ大学を訪問した際、放影研のデータ統合のニーズに対してGen3システムが良い解決策になるかどうかを議論したことを受け、生物科学部研究情報科学最高責任者のRobert Grossman博士が2日間来所した。Robert Grossman博士は、講演（「Gen3を使ったデータセンターの構築」）を行い、来所中のアーキビスト、ITD職員、放影研の主導陣と会議を開いた。会議終了後、Gen3システムを使用したパイロットプログラムを放影研で実施することが決定した。パイロット版を2021年1月まで運用し、終了した。Gen3は複雑な（そして「ベータ」ソフトウェアサイクルのため、変化し続ける）システムであった。これに加え、システムに携わるのが非常勤の学生であるということがあり、うまく行かなかった。Grossman博士との最終面談で、同博士は、当該システムはシカゴ大学のサーバーでその技術者で運用するのが最適であるため、「外部で」のインストールは非常に困難だろうとの見解に同意した。

4. Christy Henshaw博士（ウェルカムトラスト）の来所、2019年3月

デジタルプロダクションマネージャーのHenshaw博士は、放影研で「ウェルカムコレクションにおける電子化デジタル化の取組み」と題した講演を行い、ウェルカムトラストのオンラインプレゼンスに関する戦略や業務上の技術的な詳細について語った。Henshaw博士は、放影研の図書資料課職員や、講演会に招待された地元のアーキビスト達とも会議を開いた。ウェルカムトラストが使用しているツールの多くはオープンソースであり、放影研のRRC設立に役立つかもしれない共同プロジェクトについて協議した。

5. RRC支援用のサーバーの購入

1月の会議でパイロットプロジェクトについて合意に至った後、技術的要件についての議論が始まった。Gen3チームは、最低でも少なくとも2台の高性能サーバーを推奨した。放影研は3台のサーバーを購入することを決定した（Gen3用に2台、検索やアーカイブを行うウェブサーバー支援用に1台）。夏に用度課が公開入札を開始した。2019年9月にコンピューターの購入、納品が完了した。

6. Robert Grossman博士と技術者2名の来所

2019年11月、シカゴ大学のGrossman博士と技術者2名が、新規購入したサーバーに「Gen3 データコモンズ」のパイロット版をインストールするために来所した。Grossman博士は2日間、技術者は1週間滞在した。放影研で最高のシステム技術者がシカゴ大学の技術者と共に、使用するインストールツールや必要なネットワーク設定について理解するため1週間かけて共同作業を行った。インストールの初期作業は1週間で完了した。このシステムに関し、

放影研は証明書に基づく第三者認証システム（Microsoft Azureインフラ）を用いた最新の認証システムを採用する必要があった。

7. ヒトゲノムの共同研究のために、ビデオ監視および生体認証による入退室管理システム付きの安全なコンピューター室を2部屋設置
放影研ではヒトゲノムデータの導入取得を進めているが、そのための所内での計算機能力は十分ではない。この種の研究を行うためには、スーパーコンピューター（当初の目標は東北メディカル・メガバンクのスーパーコンピューター）にアクセスする必要がある。このコンピューターを使用するためには、認可された物理的スペースとコンピューターインフラが必要であった。建設業者を使い仕様書に基づき、現在放影研には、この特殊化した機能のために専用の部屋が2室設置されている。
8. オープンライブラリーサービスに関する会議や講演、日本におけるオープンデータ・イニシアチブ、オープンレポジトリに関する会議、専門家養成に関する講演講義等、会議や講義に参加した。
9. 白書の完成
10. 「デジタル化小委員会」（小野ITD部長、丸茂図書資料課長が指揮する10名の委員会）、「データ管理小委員会」（坂田疫学副部長が指揮する13名の委員会）の活動終了。
この2つの小委員会は、放影研で保存されている全ての資料の目録を作成し、デジタル化が必要な資料を特定するという作業を完了した。データ管理小委員会は全研究部と会議や聞き取りを何度も行い、20種類の資料（デジタルファイル、ノート、地図、コード化マニュアル、紙のデータカード、研究室の写真など）のリストを特定、整理した。何百万枚もの紙資料、デジタル記録、フォルダーで一杯のファイリングキャビネットや箱などに保存された資料等である。デジタル化小委員会は、データ管理小委員会と協力して、スキャンが必要な資料を特定した。
11. カルテのスキャン作業に関するパイロットプロジェクトを完了：
外部の業者と契約を結び、広島研究所臨床研究部に保管されている紙のカルテをスキャンするために必要な労力を詳しく調査するため、パイロットプロジェクトが実施された。3週間の試行期間中に、様々なワークフロー方法、オーバーヘッドスキャナーとフラットベッドスキャナーの比較、様々な劣化状況にある留め具の取り扱い方法、サイズが不揃いの紙の取り扱いなどをテストし、その後報告書が作成され、放影研に提出された。パイロットプロジェクトにより、カルテには1,000万枚の紙が使われていると推定された。業者は、カルテの資料を全てデジタル化するには約500万ドルかかると結論づけた。25人のチームであれば、3年で当該業務を完了できる。25人より少ない人数でも同じ費用でデジタル化できるが、作業期間は長くなる。長崎研究所の資料をスキャンすると、総費用が約4割増えると推定されている。

臨床研究部で確認された紙資源に加え、他の研究部でも800万から1,000万枚の紙資源が存在すると推計される。このため、総枚数はおよそ2,000万枚になる。1枚あたり約0.5ドルとして、放影研の全紙資産を外部業者が（契約ベースの）デジタル化した場合、費用は合計1,000万ドルと見積もられている。