



NATIONAL LEADERSHIP GRANTS FOR LIBRARIES

Sample Application LG-259500-OLS-26
Project Category: Applied Research

University of Maryland, School of Information

Amount awarded by IMLS:	\$171,111
Amount of cost share:	\$0

Then University of Maryland, College of Information Studies will develop and evaluate a phased Generative Artificial Intelligence (Gen-AI)-powered knowledge system that progressively extracts layered metadata, preserves provenance, links every extracted fact back to its source page, and supports a public-facing conversational interface for searching and analyzing the Franklin Delano Roosevelt Gift collection, correspondence with ordinary Americans during the Great Depression and World War II. The intended results include: a functioning GenAI powered public access system with page-level archival source links; a structured database from knowledge graph connecting people, places, gifts, topics, and sentiments; a replicable agentic GenAI workflow for use by other libraries and archives; and measurable impact through community engagement, feedback, and implementation. The beneficiaries are the public, educators, students, researchers and historians.

Attached are the following components excerpted from the original application.

- Narrative
- Schedule of completion

When preparing an application for the next deadline, be sure to follow the instructions in the most recent Notice of Funding Opportunity for the grant program to which you are applying.

CONNECT FDR:

Unlocking Franklin D. Roosevelt's Correspondence with Ordinary Americans Using GenAI

1. Introduction

The U. of Maryland is submitting a 2-year Applied Research proposal to the IMLS National Leadership Grant for Libraries (NLG-L) program for \$171,111. We propose an innovative applied research project called CONNECT FDR, that harnesses the potential of GenAI to unlock innovative historical presidential communications. The project focuses on historical education and archival practices that relate to our shared American history and aligns with NLG-L Goal 1 & Objective 1.3. GOAL 1: *'Develop, enhance, or disseminate replicable practices, programs, models, or tools to strengthen library and archival services for the American public.'* OBJECTIVE 1.3: *'Provide broad access to and preservation of information and collections through libraries and archives.'* We intend to deliver a functioning AI-powered public access system for the FDR Presidential Library's President's correspondence with ordinary Americans, a structured and provenance-preserving knowledge graph connecting people, places, gifts, topics, and sentiments, and a documented, replicable agentic GenAI workflow for processing complex archival PDFs. This project will produce scalable and generalizable methods for GenAI-enabled archival stewardship for potential use in other libraries and archives nationwide.

Details:

This project addresses a pressing national opportunity: how to provide broad, meaningful, and explainable public access to complex presidential archival collections using responsible artificial intelligence. Specifically, the project focuses on the President's Personal File 9: Gifts (PPF 9), a digitized yet structurally complex body of correspondence between Franklin D. Roosevelt and ordinary Americans during the Great Depression and World War II.¹ While these materials are preserved in their original archival order within his Presidential Library in Hyde Park, access remains limited to manual page-by-page review. This project will develop and evaluate a phased, Generative Artificial Intelligence, or GenAI-powered knowledge system that progressively extracts layered metadata, preserves provenance, links every extracted fact back to its source page, and supports a public-facing conversational interface for searching and analyzing the collection. The intended results include: (1) a functioning GenAI-powered public access system with page-level archival source links; (2) a structured database from knowledge graph connecting people, places, gifts, topics, and sentiments; (3) a replicable, documented agentic GenAI workflow for use by other libraries and archives; and (4) measurable impact through community engagement, feedback elicitation and implementation.

This project uses GenAI in two complementary ways. First, we will implement an agentic GenAI workflow² behind the scenes. In this approach, GenAI systems follow structured, multi-step processes to review each digitized folder page-by-page, identify document types, extract information in layers, connect related people and topics, and progressively refine earlier results as new information is discovered. This staged process allows the system to organize complex archival PDF pages while preserving original order and provenance. Second, we will use GenAI models to power an online public-facing interface, enabling users to ask questions in plain language and receive clear, page-level archival source links grounded directly in the original archival pages. Together, these methods transform static scans into an explainable and accessible knowledge system.

As an Applied Research project, it is guided by the following research questions explained in detail in further sections:

¹ Franklin D. Roosevelt, Papers as President: The President's Personal File, Part 1: PPF 1-500, 1933-1945 | Franklin D. Roosevelt Presidential Library & Museum, <http://www.fdrlibrary.marist.edu/archives/collections/franklin/index.php?p=collections/findingaid&id=553>

² Generative vs Agentic AI: Shaping the Future of AI Collaboration, IBM, https://www.youtube.com/watch?v=EDb37y_MhRw&t=3s

1. (RQ1) How can an agentic GenAI workflow progressively extract layered metadata from complex archival PDFs while preserving original order and provenance?
2. (RQ2) How can structured knowledge graphs enhance public access to presidential communications without replacing archival context?
3. (RQ3) How can GenAI systems be designed to self-update and retroactively improve previously processed archival data?
4. (RQ4) How do GenAI-enhanced access systems change how researchers and the public engage with historical records?

2. Project Justification

(A) The need, challenge, or opportunity and how it was identified

The Franklin D. Roosevelt Presidential Library holds one of the most significant bodies of presidential communications with ordinary Americans, created at unprecedented scale during the Depression and World War II. Within these holdings, the President's Personal File 9: Gifts (PPF 9) offers an unusually intimate and revealing record of the public's relationship with the President and his office such as people sending gifts, books, handmade items, and letters that express gratitude, concern, anxiety, civic pride, and appeals for action. This proposal addresses a clear and pressing challenge that emerged as soon as PPF 9 became digitally accessible at scale: although the collection is now available as multi-page PDF folder scans that preserve original order, meaningful access still largely requires manual page-by-page review, because each folder contains interleaved materials such as original gift letters, White House acknowledgment copies, routing and disposition notes, and narrative cross-reference annotations that point to related items elsewhere in the filing system. We estimate close to 104,000 physical pages, organized into 122 boxes. Each box contains a dozen or so folders (varies), with around 850 pages per box. Two-thirds of the box folders (currently 694) have been scanned into PDF files.

This challenge was identified through ongoing collaboration among UMD iSchool researchers and FDR Library leaders as they examined how researchers actually navigate PPF 9 and why conventional search and basic OCR are insufficient for many historical questions. The UMD team has been working with FDR Library collections since 2021 and recently published a **book chapter** on applying AI to two of their collections: the Henry Morgenthau Jr. Holocaust Project and the PPF 9 Gift File.³ In particular, PPF 9 is evidence-rich but not keyword-friendly: topics central to modern scholarship may appear indirectly or in historically specific language, meaning that relying on present-day terminology can miss relevant records (for example, "Holocaust" would not typically appear as a term in contemporary letters to President Roosevelt). The opportunity, therefore, is not simply "digitization," but the creation of an explainable, value-added access layer that preserves provenance while enabling the public and scholars to explore themes such as civic engagement, national unity during crisis, and Americans' concerns about global conflict, including early signals and evolving awareness related to Nazi persecution and the "Jewish question" in gifts such as books sent to the White House.

Finally, the timing is urgent and nationally relevant: the FY26 IMLS priorities emphasize historical education, improving visitor experiences with new technologies including GenAI, and meaningful commemoration of the 250th anniversary of American independence, with an explicit call for projects that strengthen civil society and support respectful civic dialogue. PPF 9 is a uniquely appropriate collection for these priorities because it documents Americans communicating across geography, class, and circumstance with their President offering a concrete, human-scaled pathway for public education about democratic engagement and national identity in a time of crisis.

³ BOOK CHAPTER: "From Machine Learning to Generative AI: New Computational Strategies to Analyze Public and Administration Sentiments related to the Holocaust in Presidential Library Archives", Kirsten Strigel Carter, Linde Brocato, Melanie Belkin, Rajesh Kumar Gnanasekaran, Richard Marciano. Accepted for publication on Jan. 16, 2026, in the Proceedings of the 2021 Virtual Conference on the Holocaust, hosted by the FDR Library.

(B) How this project differs from, complements, builds upon, or adapts existing models, standards, theories, scholarship, or practice

This proposal is an organic expansion of proven computational archival science work already undertaken by the project team and partners.⁴ Prior research demonstrated that large, complex archival collections often contain sophisticated original information structures created by recordkeepers and that computational methods can reconstruct and extend those structures without replacing them. Our 2026 book chapter referenced in footnote 3, demonstrates this through machine learning reconstruction of the Morgenthau Diaries indexing and validation workflows, including the development of tools and verification interfaces to support trustworthy computational access at scale. That work established a foundation: when historical records management systems are recognized and modeled computationally, the result is more accurate discovery, better context retention, and stronger alignment with archival principles than approaches that treat documents as disconnected pages of text. CONNECT FDR builds on this approach to a different but complementary corpus: instead of internal government documentation, PPF 9 captures the public’s voice in direct correspondence with the White House. The project also expands beyond prior “*AI on digitized content*” efforts in three important ways that respond to a documented gap in current practice

- First, it moves from one-time extraction to an agentic, multi-pass workflow that progressively adds layers of metadata from basic identification (who/when/where) to deeper interpretive layers (topics, sentiments, dispositions) and connects each layer to entities and relationships in a structured repository (knowledge graph) that remains backward- and forward-compatible as the system matures.
- Second, it centers explainability and provenance: each extracted claim will link directly back to the specific page image within the source folder scan and maintain reference to the archival arrangement, honoring the fact that the original White House staff filing system remains meaningful and functional for manual retrieval and evidential context.
- Third, it advances from researcher-only prototypes to a public-facing access prototype that is deliberately designed to be usable by non-specialists and to support education and engagement, aligning with IMLS’s emphasis on broad public benefit and improved visitor experiences.

This proposal also complements and strengthens ongoing FDR Library initiatives in Holocaust studies and digital humanities. The FDR Library’s 2021 conference on “American Responses to the Holocaust: Digital Possibilities” and the subsequent effort to develop a publication from that convening demonstrate both scholarly demand and public need for digital methods that expand access, interpretation, and civic education. CONNECT FDR contributes a practical, replicable infrastructure for that broader agenda by enabling cross-collection discovery patterns and by addressing the specific interpretive challenge that Holocaust-relevant signals in public correspondence may be indirect, contextual, and linguistically subtle, requiring layered historical knowledge structures rather than simplistic keyword search.

(C) Who will benefit from the project and its results

This project is designed to deliver benefits to multiple audiences, each aligned with IMLS’s mandate for national impact.

- **The general public** will benefit from a trustworthy, explainable system that supports exploration of FDR-era presidential communications through plain-language questions and evidence-linked answers, strengthening historical understanding and civic reflection in the context of the 250th anniversary commemoration priorities.
- **Educators and students** (including K–12, undergraduate, and lifelong learners) will gain an accessible pathway into primary sources that illuminate how Americans engaged their government

⁴ “Computational Archival Science (CAS) Hub”: <https://ai-collaboratory.net/cas/>.

during the crisis, supporting classroom use, museum programming, and family-friendly public learning experiences.

- **Researchers and historians** will benefit from new computational affordances that enable analysis at scale while preserving archival context such as tracing patterns by geography, comparing sentiments over time, or identifying clusters of public concerns reflected in gifts and letters (including Holocaust-relevant books and related correspondence that can be difficult to locate through conventional approaches). Archivists and librarians nationwide will benefit from the project’s replicable workflows and documentation, which translate a complex real-world presidential archive use case into adaptable models for other institutions managing scanned PDFs with layered, partially implicit organizational systems. In short, CONNECT FDR serves both as a public-facing civic history resource and as a national testbed for responsible, explainable GenAI methods that strengthen archival access without weakening provenance, trust, or the cultural record.

3. Project Work Plan

Overview

This project will build an GenAI-powered online system that allows the public to explore Franklin D. Roosevelt’s correspondence with ordinary Americans beginning with the President’s Personal File 9: Gifts (PPF 9). Rather than waiting until the end of the project to produce a finished system, we will develop it in clear, iterative phases, releasing working versions along the way. Each phase builds upon the previous one.

The long-term outcome is:

- **A structured, explainable knowledge system** that connects people, places, gifts, topics, and sentiments found in the collection.
- **An AI-powered online interface** that allows users to search and explore the collection using natural language, voice, or image input.
- **A system that always links back** to the original digitized page and its physical archival location for transparency and trust.
- **A self-improving knowledge base** that updates earlier data when new information is discovered.

At the core of this proposal are four research questions built on our earlier book chapter assessment (see Table 1). This directly addresses the central challenge identified in this project which is how to transform structurally complex but digitized presidential correspondence into an accessible, explainable, and sustainable public resource.

Table 1. Research questions, leads, and methods.

Research Questions	Research Leads	Research Methods
(RQ1) How can an agentic GenAI workflow progressively extract layered metadata from complex archival PDFs while preserving original order and provenance?	● PI: Dr. Richard Marciano: Professor, U. of Maryland iSchool & Director of the Advanced Information Collaboratory (AIC)	● Iterative AI workflow design and testing. ● Longitudinal performance measurement across iterative passes.
(RQ2) How can structured knowledge graphs enhance public access to presidential communications without replacing archival context?	● Co-PI: Dr. Rajesh Kumar Gnanasekaran: AI Solutions Assistant Director, U. of Maryland	● Benchmark comparison using human-labeled folders. ● Usability testing with mixed methods (surveys, interviews, task-based evaluation).

(RQ3) How can GenAI systems be designed to self-update and retroactively improve previously processed archival data?	● Senior Research Collaborator: Mark Conrad: Former Digital Archivist at the National Archives and Records Administration (NARA)	● Knowledge graph construction.
(RQ4) How do GenAI-enhanced access systems change how researchers and the public engage with historical records?	● Senior Research Collaborator: Dr. Linde Brocato: Metadata Librarian at the U. Arkansas	● Structured metadata validation sampling.

- **RQ1** addresses the core technical research challenge of this project. Using PPF 9 as an experimental dataset, we will design and test a structured, multi-pass agentic GenAI workflow capable of identifying document types, extracting layered metadata, and maintaining links to the original folder structure. The goal is to process the FDR collection experimentally to evaluate how complex, image-based archival PDFs can be transformed into structured data while preserving original order and evidential integrity. Findings from this phase will inform replicable methods applicable to other archival collections with similar structural complexity.
- **RQ2** examines whether a structured database that connects people, places, gifts, and topics can improve discovery while preserving the collection’s original filing system. Rather than reorganizing the archive, the research tests whether a relationship-based metadata layer can serve as a value-added access tool. Using the FDR materials as a case study, we will evaluate how such a structure can enhance navigation and thematic exploration without altering provenance. The emphasis is on developing design principles for knowledge organization in archival contexts.
- **RQ3** investigates how experimental GenAI systems can be structured to refine earlier extractions when new entities or relationships are discovered. Many digital projects produce static outputs; this research explores methods for iterative improvement and reprocessing within a controlled experimental environment. By testing self-updating mechanisms on the FDR dataset, we aim to determine whether adaptive systems can improve accuracy over time while maintaining traceability to original documents.
- **RQ4** shifts the focus from infrastructure to user impact. Through structured testing of a research prototype interface that provides page-level archival source links, we will evaluate whether GenAI-assisted discovery tools alter how users interpret, navigate, and analyze presidential communications. The objective is to measure engagement patterns, trust, and research efficiency in a controlled beta environment. Insights from this evaluation will inform broader discussions about responsible GenAI integration in libraries and archives.

Together, these four research questions ensure that the project simultaneously advances methodological rigor, strengthens archival stewardship, and produces measurable improvements in public engagement with America’s history.

Table 2:. Timeline for the 6 major phases of the project.

Sep. 1, 2026 – Aug. 31, 2028	2026				2027												2028											
	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A				
Phase 1: Build and Validate Initial AI Workflow																												
Phase 2: Expand to Multi-Layer Metadata Extraction																												
Phase 3: Develop Public-Facing AI Access System																												
Phase 4: Develop Self-Updating (“Self-Healing”) System																												
Phase 5: Community Co-Design, Testing, and Public Beta Deployment																												
Phase 6: Wrap-up																												

3.1 Project Structure: Iterative, Progressive Phases (24 Months)

● Phase 1 (Months 1–6): Build and Validate Initial Agentic GenAI Workflow

Purpose: Develop the first working “agentic GenAI workflow” capable of processing a full scanned PPF 9 folder and extracting foundational information.

Activities:

- Design a structured GenAI workflow that:
 - Separates large folder PDFs into meaningful document groupings.
 - Identifies document types (gift letter, acknowledgment letter, abstract, memorandum).
 - Extracts foundational metadata (sender, date, gift type).
- Use two previously manually labeled folders (PPF 9 and PPF 74) as validation benchmarks.
- Compare GenAI outputs against human-labeled data.
- Adjust and refine workflow for accuracy.
- Establish secure hosting system:
 - Generate stable web links for each PDF page.
 - Ensure every extracted data point can link back to its original page and folder location.

Deliverables:

- Working Agentic GenAI workflow validated against human-labeled benchmark folders.
- Page-level archival source links and traceability system.

● Phase 2 (Months 5–9): Expand to Multi-Layer Metadata Extraction

Purpose: Move beyond basic extraction and begin building layered knowledge.

Activities:

1. Add iteration passes for GenAI powered data extraction to the workflow.
 - a. Geographic locations
 - b. Names of organizations
 - c. Book titles
 - d. Topics mentioned
 - e. Emotional tone (gratitude, concern, urgency, pride, etc.)
 - f. Gift disposition (e.g., sent to State Department, stored at Hyde Park)
2. Connect extracted data into structured relationships:
 - a. Person → Sent → Gift
 - b. Gift → Related to → Topic
 - c. Letter → Expresses → Sentiment
3. Store relationships in a knowledge graph database.
4. Maintain full traceability:
 - a. Every extracted fact will link back to:

- i. Original PDF page
- ii. Folder number
- iii. Physical archive location

Deliverables: Knowledge Graph Version 1.0 which would be searchable, explainable, and linked to source images. Stakeholders will test structured search queries at the end of this phase. The system will not overwrite history, as it will preserve the original archival structure and add new layers of access.

● **Phase 3 (Months 8–12): Develop Public-Facing GenAI Access System**

Purpose: Create an accessible online system for researchers, educators, and the public.

Activities:

- Develop GenAI-powered interface allowing:
 - Plain language queries.
 - Structured filtering.
- Ensure every response includes:
 - Source page link.
 - Folder number.
 - Page-level archival source links.

Deliverables: Users will be able to:

- Ask questions in plain English:
 - “What gifts were sent from Texas in 1944?”
 - “How did citizens express concern about Europe?”
- See structured answers with:
 - Page-level archival source linking
 - Folder numbers
- The system will not generate unsourced answers.
- Every result will trace back to the archival record.

● **Phase 4 (Months 11-16): Develop Self-Updating (“Self-Healing”) System**

Purpose: Ensure the system improves itself as new entities and relationships are discovered.

Activities:

- Design workflow logic that:
 - Detects new entities, relationships, or patterns.
 - Updates the knowledge graph automatically.
 - Re-checks previously processed folders when improvements occur.
- Schedule automated reprocessing cycles.
- Conduct validation sampling to monitor improvement over time.

Deliverables:

- Knowledge Graph Version 2.0 with automated updating capability.
- Documented improvement metrics over earlier processing rounds.

● **Phase 5 (Months 16-21): Community Co-Design, Testing, and Public Beta Deployment**

Community Involvement:

- Conduct structured usability testing by conducting workshops and sessions with:
 - Archivists, historians, educators, members of the public
- Incorporate feedback through iterative updates.

Deliverables: Public Beta GenAI Access System with explainable outputs.

● **Phase 6 (Months 22–24): Wrap-up**

Purpose: Synthesize research findings, document tested methods, and disseminate evidence-based guidance to support national learning in GenAI-enhanced archival access.

Activities:

- Document the experimentally tested agentic GenAI workflow, including:
 - Multi-pass extraction methodology
 - Knowledge graph modeling approach
 - Source-linking and provenance preservation design
 - Self-updating system evaluation results
- Produce a research-focused practitioner report describing:
 - What worked
 - What did not work
 - Performance benchmarks
 - Limitations and risks
- Develop a methodological guidance document outlining:
 - Design considerations for GenAI in archival settings
 - Conditions under which such workflows are appropriate
- Conduct two dissemination sessions (virtual or hybrid) focused on presenting research findings and lessons learned rather than promoting system adoption.
- Submit final research publications and conference presentations.

Deliverables:

- Applied Research Methods and Findings Report
- Archival GenAI Design Guidance Document
- Final Evaluation and National Impact Summary

3.2 Who Will Plan, Implement, and Manage the Project

- Lead Institution: University of Maryland (UMD)

3.3 Key Personnel

- **Dr. Richard Marciano** (PI, UMD)
 - Leads project discussions.
 - Brings prior experience in reconstructing archival indexing systems using machine learning.
- **Dr. Rajesh Kumar Gnanasekaran** (Co-PI, AI Solutions Lead, UMD)
 - Designs agentic GenAI workflows and knowledge graph architecture.
 - Leads explainability and public interface development.
- **Dr. Linde Brocato** (Senior Research Collaborator & Metadata Librarian at U. Arkansas)
 - Provides historical framing and interpretive guidance.
 - Advises on presidential communications context.
- **Mark Conrad** (Senior Research Collaborator & Former National Archives Digital Archivist)
 - Oversees archival integrity and provenance preservation.
 - Ensures GenAI workflows respect original records management structures.

3.4 Evaluation and Demonstration of Impact

- Evaluation will measure:
 - Technical Performance
 1. Accuracy compared to manually labeled benchmark folders.
 2. Precision and recall of metadata extraction across iterative processing cycles.
 3. Measured improvement in extraction quality after multi-pass refinement.
 4. Reliability of page-level archival source linking.
 - Transparency and Explainability

1. Percentage of extracted data traceable to original archival pages.
2. User understanding of source-linking features.
3. Archivist validation of provenance preservation.
- Community Engagement Impact
 1. Participation in structured testing sessions.
 2. Observed changes in user navigation and discovery patterns during beta evaluation.
 3. Survey-based measures of trust, clarity, and research usefulness.
- National Research Impact
 1. Conference presentations and peer-reviewed publications produced.
 2. Citations or references to project findings in scholarly or professional venues.
 3. Attendance and engagement in dissemination sessions focused on research findings.
 4. Evidence of professional dialogue regarding the applicability of the tested methods in other archival contexts.

3.5 Applied Research Requirements

- **Research Questions**

- **RQ1:** How can an agentic GenAI workflow progressively extract layered metadata from complex archival PDFs while preserving original order and provenance?
- **RQ2:** How can structured knowledge graphs enhance public access to presidential communications without replacing archival context?
- **RQ3:** How can GenAI systems be designed to self-update and retroactively improve previously processed archival data?
- **RQ4:** How do GenAI-enhanced access systems change how researchers and the public engage with historical records?

- **Methods**

- Iterative GenAI workflow design and testing.
- Benchmark comparison using human-labeled folders.
- Structured metadata validation sampling.
- Knowledge graph construction.
- Usability testing with mixed methods (surveys, interviews, task-based evaluation).
- Longitudinal performance measurement across iterative passes.

3.6 Institutional Review Board (IRB) requirements

- Archival materials involve historical documents only.
- IRB approval will be sought for:
 - Focus groups
 - Usability testing
 - Survey and interview data collection
 - Participant data will be anonymized.
 - All data stored securely according to UMD protocols.

3.7 Research Infrastructure: Experimental AI Workstation, API Credits, and Cloud Environment

To conduct the proposed applied research, the project requires a controlled computational environment for experimentation with complex, image-based archival PDFs and multi-pass agentic GenAI workflows. The University of Maryland will assemble a dedicated AI research workstation configured for high-performance GPU processing. This workstation will support experimental tasks such as document segmentation testing, optical character recognition validation, metadata extraction benchmarking, and iterative workflow

refinement. It will function strictly as a research and development platform. Its purpose is to enable controlled experimentation, reproducibility, and performance evaluation during the grant period.

In addition to local experimentation, the project requires limited API credits to access large language models through secure cloud providers such as OpenAI, Anthropic Claude, or Google Gemini. These services will be used in a controlled research context to evaluate extraction quality, compare model performance, test explainability mechanisms, and assess cost-efficiency across providers. Because this is an applied research study, the use of multiple APIs is necessary to generate comparative findings and methodological guidance. The requested API funds support temporary, consumption-based experimentation rather than long-term service subscriptions.

Finally, limited cloud resources are required to deploy and test a research prototype interface and to host a temporary knowledge graph database during evaluation phases. These services will allow the team to assess system responsiveness, scalability, and usability under real-world testing conditions. The cloud environment will operate strictly as a beta research platform and is not intended to establish permanent operational infrastructure. All hosting and database services will be limited to the experimental period necessary to evaluate the research questions and generate generalizable findings. These resources are necessary to test real-world performance and usability but are not intended to establish long-term operational infrastructure beyond the grant period.

4. Project Results

By the end of the 24-month project, we will produce a fully evaluated research prototype that demonstrates how an agentic GenAI workflow can process complex archival PDFs and support an AI-assisted access interface using the FDR Presidential Library's President's Personal File 9 (PPF 9) as an experimental dataset. The primary outcome is not a permanent operational system, but a rigorously tested proof of concept. The project will yield: (1) a research prototype public interface that links responses to specific archival pages; (2) an experimentally constructed, provenance-preserving knowledge graph derived from selected PPF 9 folders; and (3) a documented agentic GenAI workflow tested under controlled conditions. These components will allow us to evaluate whether layered metadata extraction, structured relationship modeling, and page-level source linking can meaningfully improve access to complex digitized archival collections while preserving original order and evidential context.

The project's principal contribution is methodological. Using the FDR materials as a structured test case, we will generate validated findings about multi-pass extraction strategies, knowledge graph modeling in archival environments, self-updating system design, and user engagement with GenAI-assisted discovery tools. All workflows, schemas, and evaluation protocols will be documented in a manner that separates the experimental design from institutional specifics, ensuring that the results are transferable to other libraries and archives managing similarly structured collections. Rather than delivering an implemented service, the project will produce research findings, design principles, performance benchmarks, and documented limitations that inform future GenAI-enabled archival initiatives nationwide.

Sustainability for this project is defined in terms of knowledge dissemination rather than ongoing system operation. Results will be shared through peer-reviewed publications, conference presentations, national webinars, practitioner-focused reports, and openly available documentation. The research prototype will serve as a demonstration model illustrating both the opportunities and constraints of applying agentic GenAI in archival settings. By clearly articulating tested methods, measured outcomes, and responsible design considerations, the project will contribute durable, generalizable guidance for libraries and archives exploring GenAI-enhanced access. In this way, the long-term impact lies not in maintaining a specific system at the FDR Library, but in advancing scalable and evidence-based approaches to GenAI-enabled archival stewardship across institutions.

Schedule of Completion

Table 3. Timeline for the 6 major phases of the project.

Sep. 1, 2026 – Aug. 31, 2028	2026				2027												2028											
	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A				
Phase 1: Build and Validate Initial AI Workflow																												
Phase 2: Expand to Multi-Layer Metadata Extraction																												
Phase 3: Develop Public-Facing AI Access System																												
Phase 4: Develop Self-Updating (“Self-Healing”) System																												
Phase 5: Community Co-Design, Testing, and Public Beta Deployment																												
Phase 6: Wrap-up																												