LAURA BUSH 21st CENTURY LIBRARIAN PROGRAM

Sample Application RE-260275-OLS-26
Project Category: Early Career Research

University of Tennessee, School of Information Sciences

Amount awarded by IMLS:	$273,336
Amount of cost share:	$0

In this Early Career Research project, Dr. Kai Li at the University of Tennessee, Knoxville's School of Information Sciences will investigate a fundamental but underexplored question for libraries: how do audiences of scholars, general readers, podcast listeners, and media reviewers engage with the academic monograph? These books have traditionally been peer evaluated through scholarly book reviews published in academic journals. However, a growing ecosystem of public engagement has also emerged through public platforms that offer a source of evidence about how scholarly knowledge circulates beyond the academy. This project will construct a large-scale, openly licensed dataset linking bibliographic metadata, scholarly journal reviews, and public attention data across various platforms, enabling librarians to integrate it into their systems and workflows.

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.

Parallel Reviews: Scholarly and Public Reception of Academic Monographs

LB21 Early Career Research Proposal | University of Tennessee, Knoxville

Introduction

The University of Tennessee, Knoxville (UTK), School of Information Sciences requests $273,336 in IMLS funds to support a two-year Early Career Research project, "The Review Gap: Analyzing the Drivers of Scholarly and Public Influence for Academic Monographs." This project, led by Dr. Kai Li, Assistant Professor in the School of Information Sciences, investigates a fundamental but underexplored question for libraries: how do diverse audiences—scholars, general readers, podcast listeners, and media reviewers—engage with the academic monograph, one of the library's core intellectual assets? This project will construct a large-scale, openly licensed dataset linking bibliographic metadata, scholarly journal reviews, and public attention data across platforms such as Goodreads, Amazon, and the New Books Network. The intended results include: (1) a high-quality, downloadable monograph impact dataset; (2) at least three peer-reviewed publications; and (3) a report, titled "*Alt-Metadata: Integrating Public Engagement Data into Library Discovery and Collection Services*," providing actionable guidance for library professionals seeking to integrate public engagement data into technical services, discovery systems, and collection development workflows. This project aligns with LB21 Program Goal 2, Objective 2.2, supporting the Project Director's long-term research agenda in scholarly communication and library metadata innovation.

Project Justification

The Need: A Blind Spot in Library and Information Science

Academic monographs remain the primary vehicle of scholarly contribution in the social sciences and humanities (SSH) (Engels et al., 2018; Zuccala et al., 2018). In addition, the impact of academic books extends to various different areas, which makes them a unique object to be evaluated. Scholarly book reviews—published in academic journals—have long served as a form of peer evaluation, but they represent only one dimension of a book's influence. In parallel, a growing ecosystem of public engagement has emerged: reader reviews on Goodreads and Amazon, author interviews on podcasts such as the New Books Network, and coverage in venues like *The New York Times Book Review*. These public traces offer a rich but largely untapped source of evidence about how scholarly knowledge circulates beyond the academy.

Yet despite their centrality and the richness of data, there are **various issues in how we can combine research impact data into a single system to systematically understand the impact of academic books from various perspectives**. For example:

- Empirical research has pointed at the incomplete coverage of books in mainstream academic databases like the Web of Science and Scopus (Giménez-Toledo et al., 2017; Sivertsen & Larsen, 2012), which often leads to the systematic biases in the evaluation of the scientific impacts of books.
- The research evaluation community has made progress in developing "altmetrics"—alternative indicators of scholarly impact based on social media mentions, news coverage, and policy citations. However, this work has focused overwhelmingly on journal articles, leaving monographs largely outside the scope of current impact measurement infrastructure (Torres-Salinas et al., 2018).
- The few existing studies of book-level impact have been limited in scale, restricted to single disciplines, or focused exclusively on citation counts (Hammarfelt, 2014; Yang et al., 2021). But more importantly, such social and broader impacts have been hardly integrated into in research evaluation infrastructure and systems to support research evaluation practices (Lahikainen, 2016).

The above gap is a key motivation of the present proposal. And we argue that this gap matters for library professionals in concrete ways. Collection development librarians make acquisition and retention decisions with limited evidence about how their communities actually engage with scholarly titles. Catalogers and metadata librarians work with bibliographic records that capture formal publication attributes but not public reception. Reference librarians guide patrons toward

resources without systematic data on which scholarly works have resonated with non-specialist audiences. As academic libraries increasingly embrace community-engaged and open-access models, the absence of robust public engagement data represents a significant barrier to evidence-based practice and a significant connection between libraries and the served communities.

The PI of this project has conducted various research related to the proposed topic. First, the PI received his Ph.D. training in scientometrics and research evaluation, focusing on various types of research objects in scientific systems, such as open-source software (Li & Yan, 2018; Li et al., 2017) and research data (Zhao et al., 2018). Second, the PI has also been working on various projects focusing on bridging the library metadata and research databases, to more systematically understand the social lives of books (Li et al., 2025; Zhou et al., 2025). These past research experience, as well as the PI's past professional experience as a cataloging librarian, helped the PI to fully equip the methodological and data skills to finish this project.

Research Questions

This project is organized around three core research questions, each designed to advance both scholarly understanding and library professional practice:

- **RQ1: What factors drive the scholarly and public influence of academic monographs?** Specifically, to what extent do book topics, publisher prestige, author institutional affiliation, and disciplinary norms predict (a) academic review coverage in scholarly journals and (b) public engagement on platforms such as Goodreads and Amazon?
- **RQ2: What is the relationship between scholarly and public reception of academic books?** Do academic review sentiment and public reader sentiment converge or diverge, and how do these patterns vary across disciplines—particularly in the SSH, where the monograph remains the primary scholarly unit?
- **RQ3: How can libraries operationalize public engagement data to enhance metadata services and collection development?** What computational methods—including natural language processing (NLP) and machine learning—are most effective for enriching bibliographic records with public impact indicators, and what are the practical barriers to adoption in library technical services?

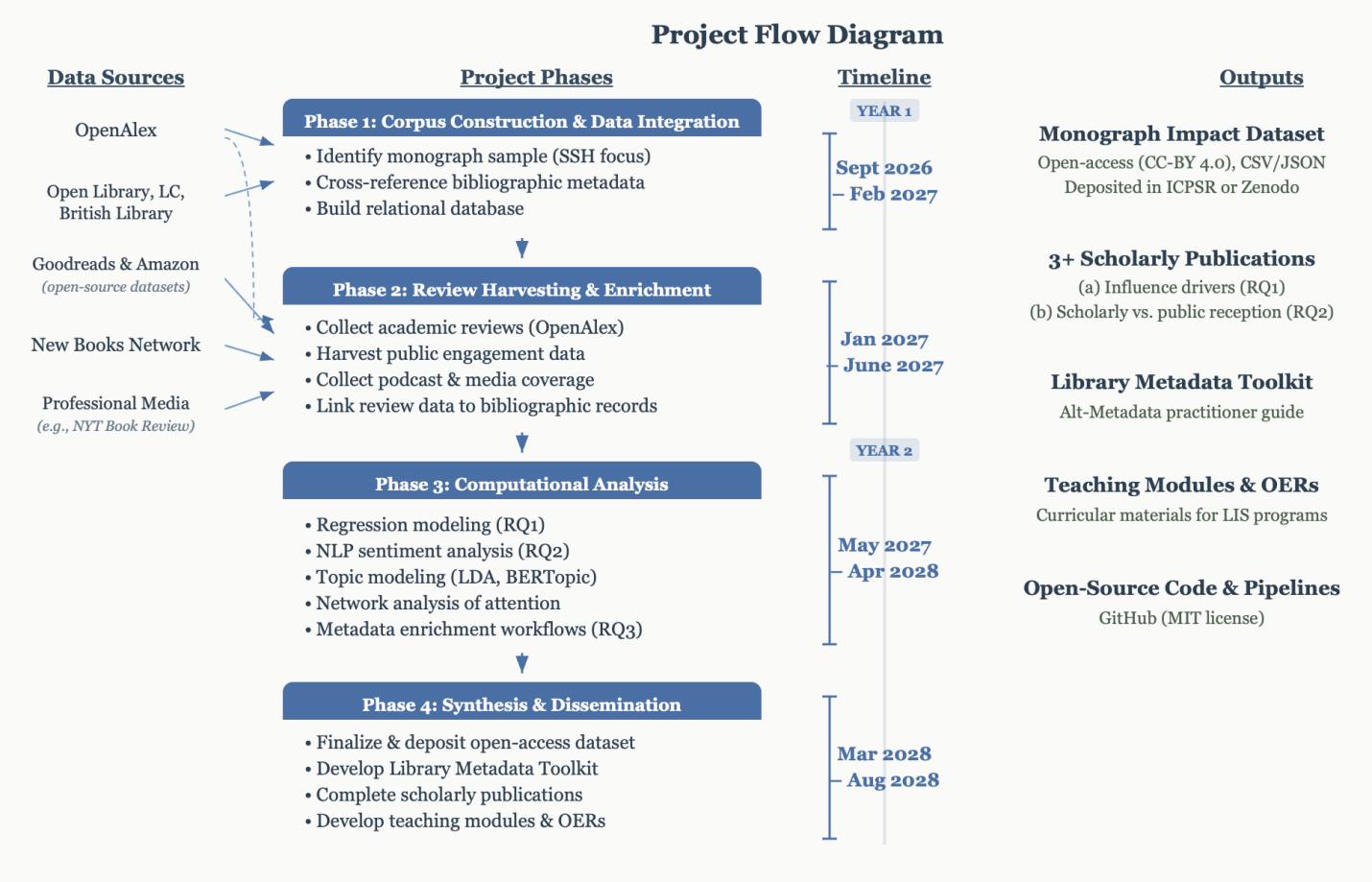
Who Will Benefit

The primary beneficiaries of this project are **library and archives professionals across the workforce spectrum**. **Practicing librarians**—particularly those in technical services, collection development, and reference—will gain evidence-based tools and methods for incorporating public engagement data into their daily work. **LIS educators and students** will benefit from new curricular content and open data resources that reflect the evolving landscape of scholarly communication. The **research evaluation and scientometrics communities** will gain a new framework for assessing monograph impact that moves beyond citation counts. Finally, the **broader scholarly community and public** will benefit from more discoverable and contextually enriched library collections.

Project Work Plan

Overview of Activities

This two-year project (September 1, 2026–August 31, 2028) proceeds in four interlocking phases, designed to build systematically from data infrastructure to analysis to practitioner-ready outputs. The overview of the timeline and outputs of the project is illustrated in the following diagram. And each phase of the project is discussed below in this section.



Phase 1: Book Corpus Construction (September 2026–February 2027). The project team will identify a broad sample of academic monographs across all disciplines, with strategic oversampling of SSH titles where the monograph is the dominant form. Bibliographic and authority metadata will be collected and cross-referenced from OpenAlex, Open Library, and open-access catalogs including the Library of Congress and the British Library. The output of this phase is a relational database linking unique book identifiers, author records, publisher information, disciplinary classifications, and publication metadata. Specifically, we will develop two corpora in different stages of the projects.

- **Corpus 1: All academic books in the New Books Network podcast**. As the first step, we will develop and test our pipeline using a small subset of academic books included in the New Books Network podcast (https://newbooksnetwork.com/), this is one of the largest academic book podcast project, covering more than 20,000 unique academic books from nearly all research fields. Using this small but representative subset, we will also be able to understand the roles played by podcasting as a means of scholarly communication.
- **Corpus 2: All academic books in OpenAlex**. Our second and final step is to parse all academic books indexed in the OpenAlex database (Priem et al., 2022). This is one of the largest academic databases in the world, with more than 477 million scholarly works indexed in the system. Given its broad coverage, it is commonly believed that this is one of the most comprehensive academic sources to understand the scholarship in SSH (Cicero & Sarlo, 2026). Currently, it includes more than 5 million academic books and 24 million book chapters, which offer a unique chance to understand the whole population of academic books and their research impacts.

Using both corpora, we are going to map academic books to a list of libraries whose catalog is openly available, including the Open Library corpus (https://openlibrary.org/developers/dumps), the Library of Congress, British Library, and Harvard Library. This is the first step to connect scholarly data and library bibliographic data, two highly relevant data sources that are traditionally siloed from each other.

Phase 2: Review Harvesting and Enrichment (January 2027–June 2027). In this project, we are going to identify the following three types of book reviews: (1) academic book reviews, (2) public book reviews, and (3) professional book reviews. We will collect these separate sources using the following methods:

- For academic book reviews, we will identify reviews mentioning a book by searching the book title in OpenAlex. Based on our preliminary testing, this approach will reach F-score of 0.9 to identify book reviews from the database.
- For public book reviews, we will use open datasets for Goodreads (https://cseweb.ucsd.edu/~jmcauley/datasets/goodreads.html) and Amazon book reviews (https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews).
- For professional book reviews, such as New York Times Book Reviews, we will use APIs provided by such venues or web scrapping (if an API is not available) to collect data from these venues.

All review and engagement data will be linked to the bibliographic records constructed in Phase 1 by mapping title, authors, and ISBN.

While the current scholarly information infrastructure makes it difficult to collect large-scale full-text academic book reviews, we strive to use a small sample (such as Corpus 1) to semi-manually collect full-text book reviews and construct text vectors of all texts related to books using advanced NLP and AI models, such as SciBERT (Beltagy et al., 2019) and Transformer (Vaswani et al., 2017). The model will be used as the input data in the next phase of analysis.

Phase 3: Computational Analysis (May 2027–April 2028). This phase addresses RQ1 and RQ2 through a multi-method computational approach.

- For RQ1, regression modeling will be used to identify the predictive factors of academic review coverage and public engagement, with independent variables including publisher prestige (operationalized through publisher rankings and imprint classifications), author prestige (operatized through the author's past productivity and research impact), author institutional affiliation (drawn from OpenAlex author records and ORCID records and operationalized through the ranking of the institution and total productivity and impact of the institution), disciplinary classification, and content outputs derived from the NLP models describe above.
- For RQ2, NLP-based sentiment analysis will be applied to both academic reviews and public reader reviews to enable systematic comparison of scholarly and public reception. Topic modeling (using BERTopic and/or results from the NLP models) will identify thematic patterns in how different audiences discuss the same works. Network analysis will map the relationships between academic and public attention to explore whether public acclaim follows, precedes, or operates independently of scholarly recognition.

Phase 4: Synthesis, Toolkit Development, and Dissemination (March 2028–August 2028). The final phase focuses on answering RQ3 and packaging research findings into outputs designed for maximum impact across both academic and practitioner communities. The open-access Monograph Impact Dataset will be finalized, documented, and deposited. The Library Metadata Toolkit will be developed as a practitioner-focused guide to integrate social data into the library metadata practice. Scholarly publications will be completed and submitted. See Project Results for full details.

While working on the previous tasks, we will also answer RQ3 by developing and testing workflows for integrating public engagement indicators into MARC and linked data bibliographic records. This applied component will evaluate which computational methods (e.g., sentiment scores, topic tags, engagement metrics) are most useful and feasible for library technical services. This phase directly explores how artificial intelligence and machine learning tools can enhance library metadata practices—investigating practical applications of these emerging technologies for the library workforce.

Project Personnel

Project Director: Kai Li, Assistant Professor, School of Information Sciences, University of Tennessee, Knoxville. He received the Ph.D. in Information Studies degree from Drexel University and has the expertise in scholarly

communication and bibliometrics, and relevant experience. The PD will devote 10% of their time to the project and will be responsible for overall project direction, research design, data analysis, and the preparation of all major deliverables.

Consultant: Dr. Yuerong Hu, Indiana University. Dr. Hu will contribute expertise in book studies. Dr. Hu will advise on methodological design, particularly collecting data from different sources and interpret results related to the lifecycle of books, and will collaborate on scholarly publications.

Graduate Research Assistant: One graduate research assistant (GRA) from the UTK School of Information Sciences will be recruited to support data collection, cleaning, and analysis activities. The GRA position will provide hands-on research training in computational methods, bibliometrics, and data management—directly contributing to the development of the next generation of LIS professionals.

Stakeholder Engagement

Stakeholder input has been incorporated into the project design from its inception. The research questions were refined through conversations with practicing metadata librarians, collection development specialists, and scholarly communication librarians at UTK and peer institutions, who identified the lack of public engagement data as a significant gap in current library practice. In the first months of the project, the Project Director will recruit a formal advisory group of 3–5 professionals, drawing from researchers and librarians at the UTK Libraries as well as scholars and practitioners at other institutions with expertise in bibliometrics, metadata services, collection development, and research evaluation. This group—complementing Dr. Hu's role as a methodological consultant—will provide ongoing feedback on research design decisions, toolkit development, and dissemination strategy, meeting virtually twice per year and reviewing draft outputs before public release. Recruiting the advisory group after the project begins, rather than pre-selecting all members, allows the Project Director to identify advisors whose expertise best matches the specific needs that emerge during the early data collection and integration phases.

Evaluation

The project will employ both formative and summative evaluation strategies. Formative evaluation will occur through regular advisory group consultations, iterative testing of NLP pipelines against manually coded samples, and mid-project review of progress against the Schedule of Completion. Summative evaluation will assess the project's success against four criteria aligned with IMLS performance measures: (1) *Effectiveness*—whether the dataset and toolkit meet the information needs identified in the project justification; (2) *Efficiency*—whether resources were used effectively to generate maximum value; (3) *Quality*—measured through advisory group assessments, peer review of publications, and user feedback on the toolkit; and (4) *Timeliness*—adherence to the project schedule. The Performance Measurement Plan provides further detail.

Project Results

Intended Results and Deliverables

This project will produce three major deliverables, each designed to address the need identified in the Project Justification and to serve audiences beyond the immediate research team:

1. The Monograph Impact Dataset. A high-quality, openly licensed (CC-BY 4.0), downloadable dataset in CSV and JSON formats containing linked metadata for thousands of academic monographs, their authors, scholarly journal reviews, and public engagement indicators across platforms. The dataset will include bibliographic fields (title, publisher, year, discipline, subject headings), author attributes (institutional affiliation, career stage where available), academic review data (source journal, review sentiment scores, review topics), and public engagement data (Goodreads ratings and review counts, Amazon ratings, New Books Network coverage, media mentions). The dataset will be deposited in a trusted domain repository (ICPSR or Zenodo) to ensure long-term preservation, discoverability, and reuse. Comprehensive documentation, a data dictionary, and a codebook will accompany the deposit.

2. Scholarly Publications. At least three peer-reviewed articles will be submitted to information science and digital humanities journals. Planned topics include: (a) the predictive factors of academic and public monograph influence (addressing RQ1); (b) disciplinary variation in the relationship between scholarly and public reception (addressing RQ2); and (c) computational methods for enriching library bibliographic records with public engagement data (addressing RQ3). All publications resulting from this project will be made available through open-access channels, in compliance with the IMLS Public Access Policy. Preprints will be deposited in an open repository (e.g., SocArXiv or the UTK institutional repository) upon submission.

3. Library Metadata Toolkit. A practitioner-focused report and accompanying implementation guide titled *Alt-Metadata: Integrating Public Engagement Data into Library Discovery and Collection Services*. The toolkit will outline step-by-step workflows for how library professionals can incorporate public impact indicators into bibliographic records, discovery systems, and collection assessment processes. It will include: recommended data sources and their strengths and limitations; sample MARC field mappings and linked data property suggestions for encoding engagement metrics; practical guidance on using NLP tools for metadata enrichment; and case examples demonstrating how libraries of different types and sizes—from research universities to community colleges and public libraries—can adapt these methods to their local contexts. The toolkit will be published as a freely downloadable PDF and archived in the UTK institutional repository. We will also organize free workshops after the publication of the report to help interested practitioners to learn more about the methods and data sources.

Adaptability and Generalizability

All three deliverables are designed for broad use across the library field. The Monograph Impact Dataset uses open data sources and standard identifiers (DOIs, ISBNs, OpenAlex IDs) that any institution can work with. The computational methods and code underlying the analysis will be released under an MIT open-source license via GitHub, enabling other researchers and library technologists to reproduce, adapt, and extend the work. The Library Metadata Toolkit specifically addresses implementation across institutional contexts, recognizing that a community college library and a large research university have different resources, systems, and patron needs. By providing scalable recommendations rather than one-size-fits-all prescriptions, the toolkit maximizes its potential for adoption across the diverse landscape of American libraries.

Dissemination Plan

Dissemination activities will extend well beyond academic publishing and will occur throughout the period of performance, not only at the project's conclusion:

- **Academic dissemination:** Presentation of methodology and preliminary findings at iConference 2028 (Year 1); presentation of full results at Association for Information Science and Technology (ASIS&T) Annual Meeting 2028 (Year 2); submission of three peer-reviewed articles to journals such as *Journal of the Association for Information Science and Technology*, *Journal of Documentation*, and *College & Research Libraries*.
- **Practitioner-facing outreach:** A freely available webinar series (2–3 sessions, hosted in partnership with the UTK Libraries) introducing the dataset, toolkit, and practical applications for working librarians, targeting audiences including metadata librarians, collection development specialists, and scholarly communication professionals. At least one workshop session at a state or regional library association conference (e.g., the Tennessee Library Association annual conference). Short-form articles or blog posts for practitioner outlets such as *Library Journal*, *The Scholarly Kitchen*, or the *ACRL Insider* blog.
- **Open infrastructure:** All code, data pipelines, NLP models, and the final dataset will be released under open licenses (MIT for code, CC-BY 4.0 for data and documentation) via GitHub and a trusted data repository. The Digital Products Plan provides additional detail on formats, standards, and sustainability.

Integration into LIS Education

A core commitment of this Early Career Research project is translating research findings into curricular resources that directly prepare the next generation of library and archives professionals. As a tenure-track faculty member with both

teaching and research responsibilities in the UTK School of Information Sciences, the PI will integrate the project's data, methods, and findings into graduate-level coursework in the following ways:

Course modules and case studies. The Monograph Impact Dataset and accompanying codebook will be developed into ready-to-use teaching modules for courses in courses like "Introduction to Data Analytics" and "Introduction to Data Visualization" being taught in the MSIS program in the UTK School of Information Sciences. These modules will provide students with hands-on experience working with real-world linked bibliographic data, conducting sentiment analysis, and interpreting public engagement metrics—skills increasingly demanded by employers across the library profession. Case studies drawn from the project's findings (e.g., comparing scholarly and public reception of monographs in a specific discipline) will be designed for use in class discussions and assignments.

Computational methods training. The NLP pipelines and data integration workflows developed during this project will serve as practical teaching tools for introducing LIS students and practitioners to computational text analysis, machine learning applications, and data science methods. As libraries increasingly seek professionals with these technical competencies, embedding project-derived exercises into the SIS curriculum directly addresses a documented skills gap in the library workforce.

Open educational resources. All teaching modules, assignment templates, and instructional guides developed from this project will be made freely available as open educational resources (OERs) through the UTK institutional repository and the project's GitHub site, enabling LIS educators at other institutions to adopt and adapt them for their own programs. This extends the project's workforce development impact well beyond a single institution, contributing to the national capacity of LIS education to train professionals equipped for data-rich, technology-enhanced library environments.

Graduate research training. The graduate research assistant funded by this project will receive intensive mentorship in computational research methods, data management, and scholarly publishing—experiences that are formative for students considering research-oriented or data-focused careers in libraries and information science. The GRA will be encouraged to co-present findings at professional conferences and to develop independent research extending the project's themes.

Sustainability

The benefits of this project are designed to persist well beyond the period of performance. The Monograph Impact Dataset, deposited in a trusted repository with comprehensive documentation, will remain freely available for reuse and extension by other researchers and library professionals. The open-source codebase will enable other institutions to replicate the data collection and enrichment pipelines with updated data. The Library Metadata Toolkit, as a freely downloadable document archived in the UTK institutional repository, will remain accessible indefinitely. The Project Director's ongoing research program in scholarly communication and library metadata ensures that the intellectual contributions of this project will continue to be developed and refined in future work. The relationships built with the advisory group and practitioner communities during this project will provide a foundation for future collaborative efforts to integrate public engagement data into library infrastructure at a national scale.

Connection to Future Work

This Early Career Research project is a foundational component of the Project Director's long-term research agenda investigating the roles of books in scholarly communication. The dataset, methods, and findings developed here will directly inform future research on longitudinal trends in monograph reception, cross-national comparisons of book impact, and the development of automated tools for real-time metadata enrichment in library systems. Collectively, these lines of inquiry advance the Project Director's ultimate goal: demonstrating the value of library metadata by making it better connected to other information sources—including social media data and scholarly publication records—so that library metadata becomes more useful to research communities interested in using quantitative and computational methods to understand the cultural landscape and publication markets.

This project lays the groundwork for two specific future initiatives:

Implementation-oriented development of library metadata enrichment pipeline: Building on the prototype workflows developed under RQ3, the Project Director plans to pursue an implementation-focused project—potentially

through the LB21 National Implementation project type—to develop automated pipelines that systematically connect library bibliographic metadata with external data sources, including public engagement platforms, scholarly databases, and open knowledge graphs. Where the current project establishes the evidence base and proof of concept, the future implementation project would pilot these integrations at scale across multiple library systems, moving from research findings to operational infrastructure.

Empirical research on the full lifecycle of academic books: The Project Director also plans to develop a research proposal for the NSF Science of Science and Innovation Policy (SciSIP) program to investigate the full lifecycle of academic books across various communities—from scholarly production and peer review through library acquisition, public readership, media coverage, and long-term cultural influence—using the quantitative and computational methods refined in the current project. Where the present study focuses on the review and reception stage, the future NSF project would expand the analytical frame to encompass the entire trajectory of how books move through and between scholarly and public spheres, contributing to the science of science's understanding of knowledge dissemination beyond journal-centric models.

Statement of National Significance

Academic monographs sit at the center of humanities and social science library collections, yet the profession currently has no systematic way of knowing how these works are received beyond the academy. Scholarly book reviews and public reader responses on platforms like Goodreads exist in parallel — two reception ecosystems for the same objects that libraries acquire, catalog, and preserve — but only one has been visible to the field. This gap leaves librarians without the evidence they need to make informed decisions about collection development, discovery services, and demonstrating the public value of their collections.

This project provides **the first large-scale empirical framework for understanding both scholarly and public engagement with academic books**. The resulting open dataset, computational tools, and practitioner toolkit directly support the training and professional development of the library workforce in emerging methods — including natural language processing, sentiment analysis, and machine learning — that are necessary for working with the unstructured, large-scale public engagement data that traditional cataloging and metadata standards were never designed to capture.

The project's findings **carry particular significance for the technical and metadata services professionals who form the backbone of library discovery and access**. As linked data standards such as BIBFRAME reshape how libraries describe and connect resources, there is a growing need — and a growing opportunity — to enrich bibliographic records with evidence of public engagement. The practitioner toolkit developed through this project provides concrete workflows for incorporating reader response signals into metadata enrichment, collection assessment rubrics, and acquisitions decision-making. These are capacities that technical services librarians currently lack not for want of interest, but for want of standardized methods and training. By building these methods openly and documenting them for professional audiences, the project directly expands the analytical toolkit available to catalogers, metadata librarians, and collection development specialists nationwide.

Critically, the project's outputs are designed for broad adoption. The open dataset, hosted on Zenodo under a Creative Commons license, and the modular Python codebase require no proprietary software or institutional infrastructure beyond a standard computer. The toolkit's workflows are usable by any academic or public library, not only research-intensive institutions. By making visible a dimension of scholarly impact that has remained largely invisible to the profession, this research equips libraries across the nation to more fully demonstrate how the collections they steward serve both scholarly communities and the broader American public.

Schedule of Completion

Year 1: September 2026 – August 2027

Activity	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Phase 1: Corpus Construction & Data Integration												
Define monograph sample criteria and scope	X	X										
Collect bibliographic metadata (OpenAlex, Open Library, LC, BL)		X	X	X	X							
Build and validate relational database				X	X	X						
Phase 2: Review Harvesting & Enrichment												
Collect academic book reviews (OpenAlex)					X	X	X	X				
Integrate Goodreads and Amazon open-source datasets					X	X	X	X	X			
Collect New Books Network and media coverage data						X	X	X	X	X		
Link review and engagement data to bibliographic records								X	X	X		
Phase 3: Computational Analysis (begins)												
Develop and validate NLP pipelines (sentiment, topic modeling)							X	X	X	X	X	X
Begin regression modeling (RQ1)									X	X	X	X
Project Management & Dissemination												
Recruit and convene advisory group	X	X	X									
Advisory group meeting #1				X								
Advisory group meeting #2										X		
GRA recruitment and onboarding	X	X										
Present methodology at ASIS&T 2027											X	
Submit interim performance report to IMLS												X

Schedule of Completion

Year 2: September 2027 – August 2028

Activity	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Phase 3: Computational Analysis (continued)												
Complete regression modeling (RQ1: influence drivers)	X	X	X									
Comparative sentiment analysis (RQ2: scholarly vs. public)	X	X	X	X								
Network analysis of attention patterns			X	X	X							
Develop prototype metadata enrichment workflows (RQ3)				X	X	X	X					
Phase 4: Synthesis & Dissemination												
Finalize and document Monograph Impact Dataset					X	X	X					
Deposit dataset in Zenodo							X	X				
Develop Library Metadata Toolkit					X	X	X	X				
Develop teaching modules and OERs						X	X	X	X			
Release code repository (GitHub) with documentation							X	X				
Draft and submit scholarly publications (3 articles)			X	X	X	X	X	X	X	X		
Deposit preprints (SocArXiv/TRACE)							X	X	X	X		
Project Management & Dissemination												
Advisory group meeting #3			X									
Advisory group meeting #4								X				
Practitioner webinar series (2–3 sessions)								X	X	X		
Present findings at ALA 2028 and/or DH2028										X	X	
State/regional library association workshop									X	X	X	
Submit final performance report to IMLS												X