

Community Data Curation Competency: Curriculum Development for Master's Education

A project team from Indiana University-Indianapolis (IUPUI) requests \$149,900 for a two-year LB21 Planning Grant to **enhance the training of Library and Information Science professionals** through creating a pilot community data curation curriculum program that builds information professionals' expertise and contributes to the communities' well-being. This project bridges the gap in existing data curation education by creating a community data curation competencies framework, developing a pilot curriculum based on this framework, and evaluating and disseminating the developed curriculum.

1. PROJECT JUSTIFICATION

1.1 Background and Project Rationale

As data becomes more pervasive in all sectors of society, there is a growing need for information professionals who can assist data curation efforts outside the academic setting. Both public and private sector community organizations (e.g., local government, non-profit organizations, community-based organizations) have become more data-driven by generating and engaging in data work (Kassen, 2013; Yoon et al., 2018). These organizations generate and use internal organizational data, use existing open data, and collect additional local data for decision-making, program development, and assessment (Yoon & Copeland, 2019). With increased data utilization, community organizations face data curation challenges without a curation expert within their organizations (Bertot et al., 2014).

Library and Information Science (LIS) professionals have spearheaded data curation, providing services and education in research data management (RDM) and scientific data curation (Varvel Jr. et al., 2012), and much of LIS data curation education has focused on academic workforce needs (Palmer et al., 2014). For example, [RDMLA](#) and [MANTRA](#) provide free self-paced online professional development curriculum on RDM for LIS professionals. [DataONE](#) provides on-demand webinars, training workshops, and open educational resources regarding scientific data management. Several organizations offer professional services, case studies, and guidance related to RDM freely online (e.g., [Digital Curation Centre](#), [Data Curation Network](#), etc.). Lastly, a free course in [Research Data Management and Sharing](#) is available through Coursera.

Recently LIS professionals have recognized the need to expand the scope of data curation curriculum and have created complimentary data-related educational programs. For example, LIS educators have extended data-related education through the increased inclusion of data science and data analytics within LIS curriculum (Murillo & Jones, 2018; Ortiz-Repiso et al., 2018; Palmer et al., 2018). Additionally, computer and data science professional development programs geared toward LIS professionals have been created. For example, [Library Carpentry](#) provides online and in-person workshops on best practices in software development (i.e., UNIX Shell, Git) and data-related skillsets (i.e., OpenRefine, TidyData). Library Carpentry materials are freely available on GitHub through a CC BY license and can be used by both LIS professionals and educators.

While these programs have provided excellent data-related curriculum resources for LIS educators and professionals, few of these educational efforts have considered the context of community-related data, and mainly focus on civic and government data. As such, LIS professionals have moderately engaged with community data challenges through eGovernment (Bertot et al., 2013), citizen science (Ignat et al., 2018), disaster preparedness (Mandel et al., 2010), and open data (Hill, 2014). Recent research reported that most librarians do not have the expertise in competencies related to civic data, calling the need to develop new MLIS curriculum that prepares students for emerging specialist data roles for working with open government data (Chaar-Pérez et al., 2021; Xiao et al., 2018). As a response, several LIS educational initiatives have brought together library professionals, government partners, LIS students and educators (e.g., [Open Data Literacy](#), [Leveraging Open Data](#), [Open Data for Public Good](#), [Building Civic Open Data Capacity through Instructional Design](#), and [mCODE](#)). Particularly, the Open Data Literacy project has aimed to create MLIS curriculum, field experiences, and open educational resources to train LIS professionals in open data literacy, targeting the public sector's need to hire data-savvy informational professionals (Palmer et al., 2021; Weber et al., 2018). Additionally, the Building Civic Open Data Capacity project aims to create MLIS curriculum to position librarians as active players in their civic data ecosystem and prepare MLIS students for positions outside of traditional library settings that involve work with civic data.

While these efforts have provided some curricular resources to educate LIS students in civic and open data literacy, an essential skill in data curation, there is still a gap as these existing efforts have not specifically addressed data curation skills and education in the context of community data. We broadly define "community data" as the data that describe the local context and are used for community decision-making. Community data may include existing open data that are freely available with minimal restriction, such as open government data and school data,

and data from private sectors and other local organizations working for community or social development. Therefore, efforts to understand the data needs of community organizations and create curricular resources based on those needs should be conducted to prepare LIS professionals for this emerging data curation context, which we name *community data curation*. As noted, many community-based organizations work with non-research data (e.g., administrative data, evaluation data, residential data, local open data), and skills to keep data meaningful and accessible, manage and preserve data for long-term use, and appraise data for the fitness of use, are critical to support their data-driven community projects. This project, thus, fills the gaps in existing efforts and complements several efforts supported by the IMLS by addressing community data curation needs and developing an MLIS pilot curriculum, which builds upon previous LIS educational achievements in research data curation and open and civic data literacy.

Our goal is to strategically partner with relevant stakeholders of community data, including community organizations, data curation practitioners, and data curation education experts, to co-create a community data curation pilot curriculum. Community-based approaches are not new in LIS as presented in recent projects (i.e., [Integrated Community-Based Learning](#)), which is creating a Community Engagement post-baccalaureate certificate. Community-based approaches strive to be intentionally and mutually beneficial to all partners. To ensure this intentionality, our project adopts a Community-Partnered Participatory Research (CPPR) approach. The CPPR approach will ensure that all critical stakeholders are engaged throughout the development of the curriculum (Jones, 2009). By adopting this CPPR approach and engaging these critical stakeholders, we will design a data curation curriculum that reflects the voices of community organizations and accounts for competencies and skills gaps in the current data curation curriculum to advance data curation education for information professionals.

1.2 Preliminary Research

Preliminary research conducted by the two PIs has indicated the need for further research and curriculum development for Community Data Curation. ***This preliminary research examined 1) the data curation needs of community agencies and 2) current Master's-level data curation education.***

Regarding the *data curation needs of community agencies*, the Co-PI (Yoon) investigated the community data curation challenges through interviews with 45 community agencies as part of the IMLS funded project ([LG-96-17-0184-17](#)). This study revealed the unique community data problems (e.g., ethics, policy, resources, and practices) and a need to train information professionals for communities' distinctive data needs (Yoon et al., 2018; Yoon & Copeland, 2019, 2020). Regarding *Masters' Data Curation Education*, the two PIs collected and analyzed data curation course syllabi from ALA-accredited MLIS programs (publication in progress). We reviewed 61 ALA Accredited MLIS programs; there were 29 courses directly related to data curation and 15 syllabi available. Through topic modeling analysis of course topics and course objectives listed in the syllabi, we found that RDM, data curation, data services, and scientific data management are the dominant topics in existing data curation MLIS courses. While there were some efforts to include user and community needs, those referred to the academic and scientific community context and did not consider the local community as potential stakeholders. This study reveals a gap in existing data curation educational efforts and indicates the need to design a curriculum that can meet community data curation needs.

1.3 IMLS Goals' Alignment

This project meets the IMLS agency-level goal of ***Championing Lifelong Learning*** by *supporting the training and professional development of the museum and library workforce*. This project will accomplish meeting this IMLS agency-level goal by 1) identifying community data curation competencies, 2) creating a community data curation pilot curriculum, and 3) evaluating and disseminating the community data curation pilot curriculum. This project meets the IMLS Laura Bush 21st Century Librarian program-level goal 3 of ***enhancing the training and professional development of the LIS workforce to meet the needs of their communities***. Specifically, this project meets objective 3.2 ***by creating and refining MLIS curriculum to build the LIS workforce skills and expertise and meet the specific needs of community organizations***. This two-year ***planning project*** will bring together the relevant stakeholders to build and expand upon the current data curation curriculum. This project *explores* community organizations' data curation needs to create community data curation competencies and *enhances* LIS educational programs by including these community data curation competencies into a pilot MLIS curriculum. This planning project will explore how LIS professionals can assist community organizations and build upon current data curation education. Furthermore, this project's evaluation model will ensure equal collaboration with community organizations, data curation practitioners, and data curation educators to secure and strengthen partnerships and relationships between LIS practitioners, educators, and community organizations.

2 PROJECT WORK PLAN

2.1 Project Goals

The *project goals* are to (1) create a community data curation competencies framework; (2) design modules to develop a community data curation pilot curriculum that incorporates the competencies to train future informational professionals; and (3) evaluate and disseminate the community data curation pilot curriculum to extend existing master's-level data curation curriculum and future planned professional development opportunities for information professionals. This project will facilitate the creation of the community data curation pilot curriculum while ensuring that all relevant stakeholders are embedded within the curriculum creation process.

The three *specific aims* that will help to achieve the overarching project goal are:

- 1) To extend the current understanding of data curation competencies to the context of community data.
- 2) To engage strategic stakeholders (educators, practitioners, and community organizations) in the process of curriculum design to reflect their specific needs.
- 3) To design a data curation pilot curriculum to fulfill the current gaps in data curation education to include community data curation competencies.

The *intended outcomes* of this project include:

- 1) **Community Data Curation 'Competencies Framework' (CF)**, a framework of data curation competencies for the community sector.
- 2) **Community Data Curation 'Pilot Curriculum' (PC)**, a pilot curriculum developed specifically for community data curation.

This project's primary intended audiences are educators, practitioners, and researchers. LIS educators will directly benefit from the creation of the CF and PC as it provides the necessary materials to modify and enhance their data curation curriculum. LIS educators can utilize the CF to determine what additional competencies need to be included in their program and course development, and can adopt and deliver the PC to their students. The CF and PC provide starting points for creating LIS community data professional development programs. Educators in data-related fields (i.e., Data Science) and fields using local and community data (i.e., Urban Planning) can utilize the CF for curriculum development. For LIS researchers and researchers in data-related fields, the initial list of existing data curation competencies and the CF will help researchers broaden their understanding of trends in the data curation field and potentially create opportunities for further research of the intersection of data and community. These research opportunities are vast, as researchers in many fields are no longer only relying on research data; researchers now also utilize data from local, community, and government organizations. Lastly, the project can serve as a model for best practices for community-engaged curriculum development and research.

This project's beneficiaries include students and community organizations. LIS students and students in related fields will benefit as they will be trained to understand community organizations' data curation needs and develop the appropriate skillsets critical to assist these organizations. In the long-term, this project will benefit community organizations, as LIS professionals and professionals in related fields will be trained to respond to their data curation needs.

2.2 Project Activities (See Supporting Document 1 – Summary of Project Activities)

Structurally this project will be conducted in three phases. *Phase one* focuses on developing the CF, *phase two* focuses on developing the PC, and *phase three* focuses on evaluating and disseminating the PC. This project's evaluation utilizes the Plan-Do-Evaluate approach from the Community-Partnered Participatory Research (CPPR) (Jones et al., 2009), ensuring that our stakeholders are equal partners throughout the project.

Phase One: Community Data Curation Competencies Framework Development (8 months: 8/2022 to 5/2023).

To achieve project goal (1), we will examine existing data curation competencies, hold focus groups with key stakeholders, and create the final Community Data Curation 'Competencies Framework' (CF).

Phase 1A: Examining of existing data curation competencies

We will first review existing data curation competencies from LIS education and professional development efforts and create an initial framework of data curation competencies. We will review, compile, and define existing data curation competencies from previous literature, professional development efforts, and grant activity to create an initial data curation competencies framework. We will collect existing literature and review the data curation competencies in that literature to compile, analyze, and operationalize those competencies. This review will involve searching LIS subject-specific databases (i.e., Library and Information Science Source), general-subject databases

(i.e., Scopus), conference proceedings, (i.e., ASIS&T, iDCC), education subject-specific databases (i.e., ERIC), and LIS education-related journals (i.e., JELIS). We will use search terms such as data curation ‘competencies,’ ‘skills,’ and ‘proficiencies’ to scope the review and focus on previously established data curation competencies. We will use citation tracking, backward and forward searching, and add additional search terms based on our knowledge, previous research, preliminary research, and our Advisory Board (AB)’s advice. Additionally, we will collect professional development documentation to review identified data curation competencies. This collection will include documentation from major professional development efforts from professional organizations, online communities, and educational efforts. Examples of where we would retrieve this documentation include data curation professional organizations (e.g., [DCC](#), [RDA](#), [RDAP](#)) and previously funded relevant data curation efforts (i.e., [RE-36-19-0081-19](#), [LG-36-19-0113-19](#)). We will manage all bibliographic items and documentation (i.e., reports, proposals) in Zotero. We will compile and collocate competencies from existing literature, professional development efforts, and previously funded projects, including definitions when available. We will conduct a content analysis to determine and remove redundancies, create initial operational definitions for each data curation competency, and create an initial community data curation competencies framework.

Phase 1B: Stakeholder focus groups to provide feedback on the initial data curation competencies framework

We will hold three 75-minute focus group sessions to review and receive feedback from relevant stakeholders on the initial community data curation competencies framework. The relevant stakeholders will include participants from 1) community organizations, 2) data curation professionals, and 3) data curation educators. We will recruit seven participants per each stakeholder group to run our focus groups (See Recruitment).

Recruitment

Community Organization Representatives will nationally be recruited by working with our AB and the Co-PI’s connection with ten nationwide community data intermediary organizations as a part of her previous work ([#LG-96-17-0184-17](#)) that have extensive data partnerships with community organizations. *Focus Group with Data Curation Professionals* will be recruited based on recommendations from our AB, through the PIs’ professional connections via direct email and professional network (i.e., RDAP, RDA). *Data Curation Educators* will be recruited from our preliminary study on Data Curation Master’s Education who previously taught relevant master’s courses, through the PI’s professional connections and professional LIS education organizations (i.e., ALISE, JESSE). To encourage the study participation, each participant will receive a \$100 gift card for their participation.

Focus Group Logistics

We will create and submit a study protocol to the IU Institutional Review Board for approval to conduct this human-subject research. The PIs will develop the focus group protocol guide, pre-test, and then conduct the 75-minute focus groups with each stakeholder group. We will use the IU Zoom platform to administer and record each focus group.

Participants will be asked to discuss their data curation needs, practices, challenges, and skills gaps (community organizations), their experiences and competencies working with various types of data and their skills gaps when working with local and community data (data curation professionals), and educational opportunities and challenges to address the identified community data curation competencies (data curation educators). All three groups will review the initial data curation competencies framework and provide feedback on which competencies are particularly pertinent to their data needs and if any additional competencies should be added to the framework. The feedback gleaned from the focus groups will be integrated into the initial community data curation competencies framework.

Upon the completion of all sessions, the research assistant will create focus group transcriptions for data analysis. The focus group transcripts will be analyzed using NVivo, a qualitative data analysis software. We will deductively analyze the transcripts based on pre-developed protocols designed around the guiding questions asked during the focus group, and capture any critical competencies specified by the stakeholders, not in the initial framework. The entire project team will code all three transcripts, compare results, discuss variations, and run intercoder reliability. Once the analysis is complete, we will update the initial Community Data Curation ‘Competencies Framework’ (CF) based on the findings and incorporate any competencies of particular importance to our stakeholder groups.

Phase 1C: Review, Update, and Creation of the Final CF

We will present the updated CF to our AB and our professional community for feedback. The updated CF will include a list of competencies, operational definitions, and any hierarchy found based on the focus group analysis.

Our AB will first asynchronously review CF through a shared document, and during *Advisory Board Meeting 1*, discuss their review feedback and suggestions. We will also present the framework at educational and professional conferences for community feedback (i.e., ALISE, RDAP). To maximize interaction and informal feedback, we will present at the poster sessions of these conferences. After collecting feedback from both our AB and community, we will make any necessary updates and create the final CF.

Phase 1D: Preparation for Phase 2

To prepare for Phase 2A (See below for the details), the research assistant will gather course, specialization, and program-level documentation from ALA-accredited MLIS programs. This documentation will include any additional data curation course syllabi that have become available since the PIs preliminary research of Masters' Data Curation curriculum. The research assistant will also examine ALA-Accredited programs for program-level goals and specializations related to data curation. A protocol will be developed similar to those protocols used by the PIs in previous curriculum research studies to gather the necessary documents (Yoon et al., 2021).

Phase 2 (Goal 2): Development of a Community Data Curation Pilot Curriculum (6/2023 to 12/2023).

To achieve our project goal (2), we will first identify competencies in the existing MLIS data curation curriculum and compare these competencies with the CF to identify gaps and determine the needed curriculum updates. We will then develop guiding instruction modules and competency-based content modules to create the Community Data Curation 'Pilot Curriculum' (PC).

Phase 2A: Identify gaps in current data curation curriculum

We will review ALA-accredited LIS school's data curation (or relevant) curriculum to identify competency gaps through analyzing the information gathered in Phase 1D. This is to update our preliminary study on Master's level data education (see p. 2. Preliminary study). While our preliminary study focused on course-level analysis, we will extend our unity of analysis to a program level (e.g., specialization, degree) to capture competencies possibly taught not in a single class. We will conduct a topic modeling of the course, specialization, and program-level learning objectives, review the identified competencies found in the current courses, specializations, and program-level learning outcomes, and identify any competency gaps by comparing the current competencies taught with the CF. From this analysis, we will develop a list of competencies that should be included in the PC.

Phase 2B: Creation of guiding instructor modules, preliminary competency-based content modules, and initial PC.

Based on the results of Phase 2A, we will develop a test course in Canvas, a learning management system, to provide both structure and easy future dissemination. The test course will consist of 1) *guiding instructor modules* and 2) *competency-based content modules*. The *guiding instructor modules* will serve as a teaching guide for instructors. Structurally, the guiding instructor modules will include a course homepage, placeholders for an instructor to personalize, and a set of pages with a course orientation. The course orientation will include an introduction to the curriculum, a history of the curriculum development, and an explanation and description of each competency found in the competency-based content modules. The guiding instructor module will include a higher-level explanation of how to use the curriculum and guidance on how to use the competency-based modules. The *competency-based content modules* will be created around each community data curation competency and will provide the curriculum content. The competency-based content modules will include teaching strategies, resources, sample lecture content, readings, and formative and summative assessments. Together the *guiding instructor modules* and *competency-based content modules* will create the initial PC.

We will work with an instructional design expert to design these asynchronous modules to ensure quality curriculum development. Additionally, as both PIs are certified [Quality Matters](#) (QM) reviewers, we will utilize our knowledge to instill the QM-based structure into the course site. QM is one of the most well-known standards for implementing best practices in online and blended courses. The QM rubric consists of 42 Specific Review Standards used to guide course development features critical to student success (i.e., learning objectives, learner interaction, accessibility) and creates ideal conditions for learning (MarylandOnline, 2020; Murillo & Jones, 2020).

Phase 2C: Review and Update of PC

For community review, we create a call for a special session at ALISE to receive feedback on the PC. Additionally, we will host *Advisory Board Meeting 2* and share the initial PC (both the guiding teaching modules and competency-based content modules) for review, comments, and suggestions. We will review the community and the AB's feedback with the instructional designer and revise and update the initial PC.

Phase 3 (Goal 3): Evaluation and Dissemination of the Community Data Curation Pilot Curriculum (1/2024 to 7/2024).

To achieve our project goal (3), we will create a Community Data Curation ‘Pilot Curriculum’ (PC) test course site, systematically evaluate the PC using the test course site, make updates based on this evaluation and community feedback, and create the final PC for dissemination.

Phase 3A: IRB and preparation for PC evaluation study

The project team will submit an IRB application to the IU Institutional Review Board to evaluate the PC. The IRB study protocol will include informed consent documentation, pre/post-test surveys, mid-curriculum surveys, end-of-curriculum surveys, and participants recruitment methods.

Phase 3B: Creation of a PC Test Course for Evaluation and Course Alignment Evaluation

The project team will create an asynchronous PC test course. The PC test course will be utilized for 1) competency-based content module evaluation and 2) course alignment evaluation. The PC test course will be built in Canvas and will include a course homepage and three initial PC competency-based content modules, among other items. Each module will contain a pretest survey, educational content, lecture materials, instruction, activities, and assessments based on the module learning objectives. The test course will include a summative assessment and a posttest survey. We will work with the instructional designer to ensure that the test course is designed to conduct the systematic evaluation with student participants that will occur in Phase 3C. We will structure the test course to ensure that student participants have enough guidance to complete the asynchronous study. For further evaluation, we will conduct an alignment analysis of the learning objectives, curriculum materials, and assessments to ensure that each course component aligns with the related competencies by utilizing the [QM Rubric](#) alignment method.

Phase 3C: PC Course Study and Evaluation

We will utilize the PC test course for evaluation with 10 MLIS student participants recruited nationwide through the MLIS blog at IUPUI, the PIs alma maters (i.e., University of North Carolina-Chapel Hill, University of Michigan, University of Iowa), our AB’s networks, and by reaching out to MLIS programs directly. We will provide MLIS students with a \$100 gift card to participate in the study. The study will take place asynchronously during May and June of 2024. We will ask participants to take the pretest survey, work through the content of three competency-based content modules, complete formative and summative assessments, and take a posttest survey within a two-week study period. We will then evaluate how well students learn the competencies through the pre/post-tests and formative/summative assessments analysis. We will examine student experiences with the curriculum through mid- and end-of-semester surveys. We will analyze the data from the pre/post-test surveys, assessments, mid/end of semester surveys using descriptive statistics in SPSS. Any open-ended survey questions will be analyzed using inductive content analysis in NVivo.

Phase 3D: Review, Update, and Creation of Final PC for Dissemination

We will host our last *Advisory Board Meeting 3* and present the evaluation results of the PC course study for final comments and review. We will then revise and finalize the PC for a final round of community review at professional conferences and journal articles, such as a JELIS. We will disseminate the product through our professional network (ALISE and JESSE listserv) for community review. The final PC will be made publicly available via Canvas Network and OERCommons. (See **Dissemination Plan**)

2.3 Evaluation Plan (See also **Performance Measurement Plan**)

Two types of evaluation, 1) research and 2) project, will be conducted to ensure this project’s effectiveness, efficiency, quality, and timeliness.

Research Evaluation: The rigor of the empirical research and research practices will be evaluated by the project team and our Advisory Board (AB) members. The project team will employ strategies to ensure the rigor of our research, including testing intercoder reliability to evaluate the coding schema agreement for the content analysis in Phase 1 (Focus Groups) and Phase 3 (Open-Ended Survey Questions). Both human research-related studies, the Focus Groups in Phase 1 and the PC Course Study in Phase 3, will undergo IRB approval to ensure ethical treatment of subjects. Additionally, all focus group and survey protocols will be piloted for review and feedback.

Throughout Phase 1, our AB members will provide feedback and recommendations for the research activities, including the literature search and focus group protocol development. Throughout Phase 2, our AB members and our instructional design colleague will provide feedback and recommendations on the competency comparison of the PC and the course, specialization, and program-level learning outcomes. Throughout Phase 2 and 3, our AB members and our instructional design colleague will provide feedback and recommendations for the research

activities related to the PC Course Study, including feedback on the pre/post-test surveys and mid/end-of-semester surveys. By applying the QM Rubric method, we will utilize a research-based method for course alignment.

Project Evaluation: We will use the Plan-Do-Evaluate approach from the Community-Partnered Participatory Research (CPPR) (Jones et al., 2009) to evaluate our processes, activities, and outcomes to integrate the AB's perspectives. The AB members will provide critical expertise, specific feedback, and suggestions throughout the project. For Phase 1, we will asynchronously share our planned activities with the AB for review and feedback (e.g., search strategies, focus group protocol, etc.) (Plan); we will then conduct the planned activities, making any changes as based on input from the AB (Do); at the end of Phase 1, we will hold *Advisory Board Meeting 1* to evaluate the activities and outcomes of Phase 1 (Evaluate). For Phase 2, during *Advisory Board Meeting 1*, we will share our planned activities for Phase 2 with our AB for review and feedback (e.g., creation of PC) (Plan); we will then conduct the planned activities, making changes based on the input from the AB (Do); at the end Phase 2, we will hold *Advisory Board Meeting 2* to evaluate the activities and outcomes of Phase 2 (Evaluate). During *Advisory Board Meeting 2*, we will share our planned activities for Phase 3 with our AB for review and feedback (e.g., evaluation of PC) (Plan); we will then conduct the planned activities, making changes based on the input from the AB (Do). Lastly, during *Advisory Board Meeting 3*, the AB will provide feedback on the final PC for final evaluation (Evaluate). Feedback from the community organizers is critical to ensure that the CF and PC are developed using an outside-in approach to include the communities' perspectives.

Additionally, we will use the **Schedule of Completion, Summary of Project Activities, and Performance Measurement Plan** to ensure appropriate project management and meeting of deliverables. We have preemptively begun activities to ensure the efficiency of outcomes (see **Preliminary Research**). For additional community feedback, we will share the project outcomes at relevant conferences (see **Dissemination Plan**). During year 1, we will present the CF at the poster session at ALISE or RDAP for community feedback and discussion. During year 2, we will have a special session at ALISE and present the PC to collect MLIS educators' input.

2.4 Project Resources

Project Team

PI: Dr. Angela P. Murillo is an Assistant Professor in the School of Informatics and Computing (SOIC) at IUPUI. Her research areas include scientific data curation, scientific data cyberinfrastructure, and data education. She is Co-PI of the CI Compass Cyberinfrastructure Center of Excellence ([NSF # 2127548](#)). She leads the research related to large-scale data archiving and is co-director of the Student Internship Program. She was the PI of an IUPUI STEM Education Innovation and Research Institute (SEIRI) Seed Grant ([SSG-2019-06](#)). She was also Senior Personnel NSF Pilot Study for a Cyberinfrastructure Center of Excellent ([NSF # 1842042](#)). She will lead all aspects of the project and serve as the Project Manager for the two years of the project, including the project, evaluation, dissemination, and other activities outlined in this proposal.

Co-PI: Dr. Ayoung Yoon is an Associate Professor in SOIC, IUPUI. Her research areas include data curation, data reuse, open data, and citizens' data engagement. She was the PI of Data Reuse for Local Community ([LG-96-17-0184-17](#)), researching community data reuse and curation and Co-PI of mCODE ([LG-250098-OLS-21](#)). She was also the PI of previous IMLS funded project, Library Capacity Assessment and Development for Big Data Curation ([LG-72-17-0139-17](#)). She will participate in all research activities in this project.

Graduate Student Research Assistant: The Graduate Student Researcher will assist with the research of data curation competencies, focus group analysis, test course creation, PC course study analysis, and project logistics.

Instructional Designer: The instructional designer will assist and provide guidance for the creation of the PC in Phase 2 and the Canvas course that will be utilized for the PC course study in Phase 3.

Advisory Board Members (AB) (See Supporting Document 2 – Letters of Commitment)

Our AB consists of three community organization partners, two data curation practitioners, two data curation educators, and one community engaged-participatory research expert.

Community organization partners:

- **Keep Indianapolis Beautiful (KIB)** is an environmental and community nonprofit with a mission to help people and nature thrive. They are an award-winning affiliate of Keep America Beautiful, Inc., a national organization dedicated to preserving the natural beauty and environment in American communities. KIB supports over 800 community projects with nearly 20,000 volunteers and was founded in 1976. Contact Person: *Kristina Uland* is the Senior Vice President of Development and External Affairs at Keep Indianapolis Beautiful, Inc. She has over 25 years of experience working with nonprofit organizations.

- **Hope Training Academy (HTA)** is an education training provider owned and operated by Video Game Palooza (VGP), a 501(c)(3) educational nonprofit dedicated to helping end the cycle of poverty by offering underemployed adults' computer-based training. Contact person: *Sarah Zike* is the Executive Director of Hope Training Academy. She has been a nonprofit professional in Indianapolis for over 15 years and brings combined expertise in business development, adult education, strategic planning, and curriculum development. She has worked with hundreds of community organizations and has expertise in community organization data needs.
- **Office of Sustainability, City of Indianapolis (ThriveIndy)** uses best practices in sustainability to enhance the quality of life for Indianapolis and Marion County residents. As a Department of Public Works division, the Office of Sustainability promotes an environmentally sustainable city through programs and initiatives. These initiatives improve air quality, stormwater management, energy efficiency, and environmental planning. Contact Person: *Mo McReynolds* is the Senior Project Manager for the Office of Sustainability, City of Indianapolis. She has experience working in local government with projects requiring complex local government datasets.

Data curation practitioners:

- **Sarah Dorphinghaus** is the director of Digital Services and the University of Kentucky Libraries Special Collections Research Center and has recent scholarship in digital curation workflows and embedding archives in pedagogical models.
- **Minglu Wang** is an Assistant Librarian and a Research Data Management Librarian at York University Libraries. She is a member of the Research Intelligence Expert Group, The Portage Network, a part of Canada's new digital research infrastructure organization, and the Digital Research Alliance of Canada.

Data curation educators:

- **Dr. Jeonghyun (Annie) Kim** is an Associate Professor at the University of North Texas. She has served as director of an IMLS-funded digital curation and data management graduate academic certificate program at the University of North Texas while researching data management and curation and LIS workforce development.
- **Dr. Nicholas Weber** is an Assistant Professor at the University of Washington School of Information. He serves as the Co-PI for the IMLS funded [ODPG](#), which incorporates open data to create curriculum materials.

Community-engaged participatory research expert:

- **Dr. Barbara Pierce** is an Associate Professor in the School of Social Work at IUPUI. She is an expert in community-engaged participatory action research and evaluation. She is a recipient of the Charles R. Bantz Community-Engaged Fellowship.

Budget (See Budget Justification)

The total budget requested from IMLS is \$149,900 (Total direct costs: \$94,574, Indirect costs (F&A @58.5%): \$55,326). The budget includes salary for the PI and Co-PI (\$40,142), one Graduate Assistant (\$28,928), consultation stipends for the Advisory Board Members (\$5,500), an incentive for focus group participants (\$2,100), an incentive for student participants (\$1,000), instructional design fees (\$3,000), software for qualitative analysis (\$300), and one domestic travel cost (\$2,500) for both PIs.

Timeline

This is a two-year project. See **Schedule of Activities** and **Supporting Document 1-Schedule of Completion**.

2.5 Communication and Dissemination Plan

The PIs will ensure effective communication between the project team and partners. The project team will schedule AB meetings at least two months prior to the meetings to ensure that members are available and can make any necessary adjustments in their scheduling. As the instructional designer will be an integral part of Phase 2 and 3, the PIs will schedule meetings with them every two weeks during those project phases. For both the AB and the Instructional Designer, the PIs will create a shared collaborative document folder where they can work asynchronously to assist with the collaboration process.

The project outcomes will be shared for review and feedback by the professional community. During Phase 1, the project team will present initial versions of the CF at professional conferences, including ALISE and RDAP. The project team will present at the poster sessions for opportunities to interact with MLIS educators and data curation practitioners. During Phase 2, we will create a call for a special session at ALISE to bring together LIS educators working with local or community organizations, community data, data curation, and open data. We will present the CF and PC for community review and feedback on the framework and curriculum during this ALISE special session. We will create a shared document for notetaking and feedback on our project's products. We will

write a JELIS article based on our curriculum development to share with the MLIS community. We will attend the Indiana Library Federation conference to present project findings and receive feedback from local LIS professionals.

The project products, including the CF, CP, conference presentations, and journal articles, will be disseminated through the project website and disseminated through professional networks (i.e., ALISE and JESSE listserv) for community reviews and feedback. The final PC will be made publicly available via the Canvas Network and OER Commons. The project team will openly disseminate project materials through ScholarWorks (IUPUI's Institutional Repository), DataWorks (IUPUI's Data Repository), and other open dissemination networks (i.e., Open Science Framework, ResearchGate).

2.6 Sustainability Plan and Future Work

The competency-based content modules from the PC will be implemented at the IUPUI MLIS program through either a special topics course or module integration into relevant courses. The PIs will utilize both the CF and the PC to continue researching and creating curricular materials with regard to community data curation. The relationships developed among the PIs, educators, and practitioners will create a larger community interested in community data curation education, research, and practice. The PIs will utilize the LIS professional connections created during the ALISE special session and the focus groups to continue cultivating this community of practice and future collaboration for research and funded projects to continue expanding upon the research and curriculum development in community data curation. Since the project outcomes will be made openly available through ScholarWorks, DataWorks, the CanvasNetwork, and OERCommons, these products will be sustained and maintained through these repositories, which will preserve the items and migrate overtime as needed. The PIs will continue to present the products at professional conferences and journal articles to ensure other LIS educators and researchers are aware of the products. Through this extensive dissemination, other LIS programs will be able to utilize the disseminated curriculum products. These curricular implementations will train LIS professionals, create leaders in community data curation, and continue cultivating the community of practice.

Planned Future Work

The PIs will utilize this project as a guide to develop a community-engaged curriculum and establish collaborative relationships with community organizations. The relationships formed between the PIs and the community organizers will provide future opportunities for collaboration beyond this project and will provide a solid foundation to expand our work with and impact on local communities. Upon completing this project, the PIs will submit an LB21 project grant to create a complete and revised Community Data Curation Curriculum based on our Pilot Curriculum study and incorporate a practicum-based experiences component for MLIS students at local community organizations. The practicum curriculum will allow students to have hands-on experiences in real settings while allowing educators to continue developing the CF and Community Data Curation Curriculum. The relationship with local community organizations that will develop during this project will be an important asset to our future projects' success, and the PIs will also expand the partnership outside of local areas to seek an opportunity for virtual practicum experiences.

3 DIVERSITY PLAN

This project will support and strengthen the LIS professional commitment to diversity, equity, and inclusion in several ways. The Project Team, including the PIs, Advisory Board (AB) members, and Graduate Student Assistant, brings together diverse backgrounds to the project, including race, ethnicity, age, subject matter expertise, professional experience, and diversity, equity, and inclusion leadership. The two PIs will leverage the [Wilma Gibbs Moore Graduate Endowed Scholarship](#), which aims to enhance diversity within the library profession, to recruit student workers for this project and target students from underrepresented backgrounds to contribute to this project. The project team will also identify organizations that serve underrepresented communities for participation in the project, with the help of our AB, IUPUI Office of Community Engagement, and the project team's national connections. By targeting organizations that serve underrepresented communities, this project will examine the needs of organizations that focus on and consider the needs of underrepresented communities and ensure diverse perspectives are represented in this project. During Phase 1, the project team will ask focus group participants to address diversity and inclusion issues relevant to data curation curriculum during the focus group discussion. During Phase 2 and 3, when creating the PC, the project team will utilize best practices and techniques for inclusive and anti-racist curriculum when developing our curriculum modules. Additionally, the project outcomes will integrate diverse perspectives during creation as the AB will be involved in all project phases due to the Plan-Do-Evaluate

approach. These outcomes, including the CF and PC, will be integrated into LIS education, ensuring that current and future MLS students will be trained and educated from this diverse perspective and will assist in establishing long-term LIS leadership in community data curation.

4 PROJECT RESULTS

This project impacts the data curation professional realm by ensuring the library and information science educators, researchers, and practitioners are equipped to respond to organizations' data curation needs. This research creates tangible products for rectifying gaps in current data curation education, research, and practice.

Education

The majority of data curation educational efforts have focused on research and scientific data management, which has produced a gap in knowledge of LIS-trained professionals in community data curation. The Community Data Curation 'Competencies Framework' (CF) and the Community Data Curation Pilot Curriculum (PC) will assist in filling this gap in knowledge. The PC will be immediately implemented into the IUPUI MLIS program through either a special topics course or existing Digital Curation (DC) specialization courses. This implementation will be logistically simple because the courses already exist and will not have to go through an approval process. And likely, many institutions could follow a similar implementation model, as special topics courses are quite common. The DC specialization is one of the growing specializations in the IUPUI MLIS program. Additionally, the IUPUI MLIS program, an online program, has a steady growth of incoming students nationwide.

As there are few existing resources for community data curation education, the CF and PC will provide the greater LIS educational community tangible resources that can be utilized to enhance or create new courses. Since these products will be structured as educational modules in Canvas, this will provide easy implementation for LIS educators. Since the PC includes both guiding instructor modules and competency-based content modules, this curriculum can be implemented by the broader LIS educator community. Additionally, since the curriculum will be made available via the CanvasNetwork and OERCommons, LIS educators and educators in other fields will have access to the curriculum.

While our efforts focus on the master's level, the impact and adoption of our products can also be expanded to practitioners for professional development programs. While LIS professionals have a long history of collaborating with community members and organizations, this tradition has not been fully integrated into LIS educational efforts, especially curriculum development, and our project can serve as an example for other educators.

Research

As data curation research has focused on research data management or developing technical infrastructure to manage data, this project will broaden the research agenda for LIS researchers and researchers in related fields by creating the CF and PC for researchers to utilize in both research and education studies. This project complements existing data curation research by integrating the perspectives of community organizations based on their data curation needs and suggests several questions to further study, such as how community data curation competencies compare to research data curation competencies.

Practice

This project will broaden the role of library and information professionals. While many data curation professionals trained from LIS education work in academic or research settings, with the new curriculum developed from this project, LIS professionals will become skilled in assisting with data curation needs and challenges in various contexts, such as non-profit, local government, and community-based organizations. With the growing demand for data curation professionals at those organizations, it is critical to broaden the roles and skillsets of LIS professionals. Further, having prepared LIS professionals for those new needs will impact local community organizations' data workflow and practices, contributing to the larger community data ecosystem.

Digital Products Plan

Type

This project will produce digital products from the compiling and content analysis of existing literature, deductive content analysis of the focus groups, video recordings of focus groups, focus group transcripts, course, specialization, and course level objectives topic models, descriptive statistics and content analysis of surveys from the evaluation study of a test course site.

These digital products will be in spreadsheet, text format, and code. Formats will include PDFs, Word docs, PowerPoint, CSV, Excel, and R. This project will also produce curricular materials in the form of a Canvas Course site.

We will submit appropriate IRB for all aspects of this project that involve human subjects and will follow the appropriate IRB procedure. All personally identifiable information (PII) and institutional names will be anonymized prior to publication. No PII will be kept within the research data and will be destroyed once this project has been completed. All focus group video recordings and transcripts will be used only for internal analysis and will be destroyed once the analysis is complete.

The digital products created and published will not include any personal information or identifying information of the organizations. Additionally, we will work with our advisory board to ensure that our digital products are culturally sensitive as needed.

Availability

The data products will be deposited in IUPUI ScholarWorks (scholarworks.iupui.edu), IUPUI DataWorks (dataworks.iupui.edu), and other open dissemination networks (i.e., Open Science Framework, Zenodo, ResearchGate). The curricular materials will be shared through Canvas Network and OERCommons.

Access

The PIs will hold the copyrights of the digital products. We will assign Creative Commons Attribution (CC By 4.0) license. This license promotes sharing, reusing, and dissemination. The users are able to share (copy and redistribute) and adapt (remix, transform, and build upon) the digital products. They must provide attribution, a link to the license, and indicate if changes were made. Lastly, they must not be used for commercial purposes. The licensing information will be included in all materials and users will be notified of the licensing information through ScholarWorks, DataWorks, Canvas Network, and OERCommons.

Sustainability

All data products will be stored during the duration of the project using IU's One Drive, which provides backup and sharing of data products. Additionally, the data products will be shared through IUPUI's institutional repositories ScholarWorks and Data Works. The IUPUI ScholarWorks and DataWorks repositories will preserve, backup, and ensure metadata is migrated over-time when needed.

Organizational Profile

Organizational Description

Indiana University is a major multi-campus public research institution, grounded in the liberal arts and sciences, and a world leader in professional, medical, and technological education. Indiana University's mission is to provide broad access to undergraduate and graduate education for students throughout Indiana, the United States, and the world, as well as outstanding academic and cultural programs and student services.

Indiana University–Purdue University Indianapolis (IUPUI), a partnership between Indiana and Purdue universities, is Indiana's urban research and academic health sciences campus. IUPUI's mission is to advance the state of Indiana and the intellectual growth of its citizens to the highest levels nationally and internationally through research and creative activity, teaching and learning, and civic engagement. By offering a distinctive range of bachelor's, master's, professional, and Ph.D. degrees, IUPUI promotes the educational, cultural, and economic development of central Indiana and beyond through innovative collaborations, external partnerships, and a strong commitment to diversity.

The Indiana University School of Informatics and Computing at IUPUI fosters a broad and interdisciplinary view of informatics and uses this view to explore and expand knowledge in informatics education and research. Along with the many schools and departments located on the Indiana University Purdue University urban Indianapolis campus, the School is firmly committed to a welcoming environment, a diverse faculty and student body, and to efforts which support Indiana's economic development.

The mission of the Indiana University School of Informatics and Computing is to excel in education, research, and civic engagement in the field of informatics, an integrative discipline which advances knowledge in computing, information, and media technologies; the implications those technologies have for individuals and society; and their application to any field of study adapting to the challenges of the Information Age.