Data Reuse for Local Community

Abstract

The research team from the Department of Library and Information Science (DLIS) at Indiana University in Indianapolis (IUPUI) will explore the practices of community data reusers, with the aim to offer insights into developing and adapting public-library services to meet community data reusers' needs.

Data possess significant potential to address current societal problems (e.g., in education, health, economic development, and the environment) not only at the federal and state levels but also in smaller communities and neighborhoods and the lives of individuals. Given this emerging community context for data reuse, public librarians can play an important role in facilitating community-level interaction with data. This project will support community members' data reuse by connecting them with existing resources and data infrastructure, with the goal to create equitable access to the vast data stores being collected about users' communities, particularly through the public libraries' engagement with data and communities. To better support communities, the project team will (1) describe community members' data-reuse practices and strategies and (2) map these practices and strategies to existing infrastructure, focusing on public libraries.

This three-year project will follow a multi-phase research design. In phase 1, the project team will conduct interviews and surveys with community data reusers to understand their data practices and reuse workflow. The results will support development of a theoretical model of community-inclusive data exchange/workflow. In phase 2, the team will conduct a study on librarians to investigate their roles in supporting their communities' data needs, efforts, services, challenges, and opportunities and the knowledge and skills librarians need to effectively deploy data services. In phase 3, the team will utilize the findings to improve or develop library services and will collaborate with partner libraries to design prototype library services.

This project will capture knowledge of community members' current data workflow to identify barriers and to propose a model for community-inclusive data workflow; advance understanding of public librarians' role and potential to support community members' data reuse; and contribute to a collaborative data-reuse culture among local data infrastructure by helping public libraries design data services to support community members' needs and by offer useful educational resources for librarians to perform such roles. Three major audiences will benefit from the project: community members (citizens) interested in using/reusing data for developing their communities and making decisions relevant to their lives; local data infrastructure and organizations that provide data services and assistance to diverse data reusers, including community members; and public librarians who want to build their capacity to meet community members' data needs.

Data Reuse for Local Community

STATEMENT OF NEED

The potential for data to help address current societal problems (e.g., education, health, economic development, and the environment) is significant not only at the federal and state levels but also in smaller communities, in neighborhoods and in the lives of individuals. Given this emerging community context for data reuse, public librarians have a potentially important role to play in facilitating community-level interaction with data. Many have argued data have value as a source for a community's economic development, disaster planning, and decision-making in citizens' daily lives (Kassen, 2013; Levin & Schneir, 2015). In the report *A World That Counts, Mobilizing the Data Revolution*, the United Nations Secretary-General's Independent Expert Advisory Group (IEAG, 2014) also underscored the significance of data, particularly for underdeveloped countries or neighborhoods, arguing that "data are the lifeblood of decision-making" and the recent data revolution brings the opportunity to design, monitor and evaluate effective policies for communities (p. 4).

This research project intends to explore the evolution of public libraries' engagement with community-based information initiatives through the lens of data reuse. Currently, public libraries serve communities by providing access to many different types of existing data resources (e.g., Census and school statistics) and have the potential to facilitate citizen engagement with data and data professionals. The Boston Public Library's recent initiative to develop an open data collection for the City of Boston demonstrates this potential (Open Data to Open Knowledge, 2017). There is a need for public librarians to connect users to data sources, as well as to the requisite training and tools that will help users make meaningful connections to data in the context of the users' everyday lives. For several years, public libraries have provided educational services to community members to develop information literacy skills. Although public librarians themselves need to develop skills in a new domain, data literacy, they are also in the position to provide data literacy education for community members to learn how to identify and evaluate relevant data for their own use and analysis, using their experiences and expertise in teaching information literacy skills. Given the interdependence of public libraries and their communities, increasing access to local data sources will allow community issues to be approached collectively through evidenced-based decision-making.

Our two major research questions in this project are as follows:

1. What are the processes, strategies, and challenges in community members' data reuse?
2. How do public librarians see their role in supporting data reuse for communities? How can the current data infrastructure become more inclusive and collaborative to support communities in data work?

RQ1. What are the processes, strategies, and challenges in community members' data reuse?

A few cases demonstrate the potential of data reuse for community members. Chicago's open data portal driven by open government data initiatives resulted in several citizen-initiated projects that developed tools for serving Chicago's diverse community to improve the transparency of the local government's political process (e.g., chicagolobbyists.org) and to ensure a certain neighborhood is maintained as a safe, open, and healthy green space for the local community (e.g., mi parquet Project; Kassen, 2013). Often, local organizations lead these projects. For example, the Green and Healthy Homes Initiative (GHHI) in Baltimore utilized health and housing data to provide healthy homes for communities and advocated for increased efforts to prevent lead poisoning (Urquilla & Shelton, 2015). These cases show how different community members (e.g., city planners, local government officers, economic developers, educational consultants from nonprofits, foundations and corporate organizations, and the public, such as parents with concerns about childcare and school systems) can utilize or reuse data collected by the private and public sectors that describe regional areas and population groups in order to better understand the factors influencing the communities.

Because of this emerging interest in using data for the community, many discussions are taking place regarding how to make the data open and sharable, which calls for action on the part of the public sector. However, little is known about community members' data practices, experiences, and perspectives, such as how they utilize data, for what purpose, and through what kind of support. A few existing projects utilizing data for a community provide a limited understanding in a big city context working with open government data to develop a tool for local residents; however, less is known about the actual process of community members utilizing the

data. If more were known, then library services could be developed to support use of the data. This project intends to understand community members' data practices and needs, and develop services accordingly.

To build more of a foundation, the PI conducted a preliminary study in 2016, funded by the Research Data Alliance (RDA)/US Data Share program through the Alfred P. Sloan Foundation (#G-2014-13746), in a small city context in the Midwest. The findings suggest that community members' data needs and practices are different from those of scientists, and community members have limited interactions with the local data infrastructure (e.g., data repositories and library data services). Instead, community members relied on regional community research centers (CRCs) that connect them with data sources, partnerships (e.g., between academia and the community), and consultations. In northeast Indiana, for example, the Community Research Institute (CRI) in Fort Wayne promotes the growth and vitality of the regional community and provides research and analytical services. The findings informed the overall study design, and we will continue to explore the data reuse practices of communities in different contexts and city sizes (e.g., Grand Rapids, MI; Milwaukee, WI; and Boston, MA). The findings from the proposed study will help us to understand community members' current data workflow, identify the barriers and hindrances community members experience, and develop a theoretical model of community-inclusive data practices among the local infrastructure to overcome the challenges and support community members' data work, with practical recommendations for public librarians to engage with a community-inclusive data workflow.

RQ2. How do public librarians see their role in supporting data reuse for communities? How can the current data infrastructure become more inclusive to support communities in data work?

Public libraries' engagement with data and communities' data work are very timely. Few public resources are available to the public despite increasing public demand, and many data are not systematically or routinely shared for reuse with communities (RDA Data for Development Interest Group, n.d.; Urquilla & Shelton, 2015). Government data are traditionally shared with citizens, but Douglass et al. (2014) pointed out that there are still structural and technical infrastructure barriers to be overcome for data to be fully accessible and easily usable. Data collected by university and research institutions are particularly not well shared with communities (RDA Data for Development Interest Group, n.d.). Research data are usually shared within scientists' own communities of practice, although researchers often share data interpretations to support community development. The IEAG (2014) pointed out that the unequal access and use of data are a barrier to enabling data to play their role in the realization of sustainable development and argued that it is critical for all stakeholders (e.g., government, regional institutions, public/private data producers, academics, and citizens) to play a role in creating a more equitable data ecosystem. Public libraries are a stakeholder that can play a role in connecting community members with existing data sources and teaching community members the necessary skills for using data, by actively engaging and collaborating with existing data infrastructure and systems.

One striking finding from the preliminary study was the lack of notable interactions between most community members and public libraries. Although the role of CRCs is critical in supporting a community's data work, there is a need for the remaining population to be served because the priority of CRCs is not always individuals (e.g., parents in the example above). Public libraries provide access to some data sources in the provision of general reference services, and public libraries have the potential to expand this current effort by facilitating citizens' engagement with data and by connecting citizens with other data infrastructures, such as CRCs. We will explore the views of public librarians on their roles, perceived barriers to those roles, and the ways in which public libraries can more actively engage with data and community members to support their data work. Using the findings from R1 and R2, we will develop a prototype library data service for users while compiling educational resources for librarians and design workshops for librarians to enhance their awareness, inform them about the significance of their roles, and help them work with community members.

PROJECT DESIGN

Project Goals and Intended Outcomes

The goal of this project is to support community members' data reuse ***by facilitating their connection with existing resources and data infrastructure in order to create equitable access to the vast stores of data being collected about their communities***. Our focus is public libraries. We define *data reuse* as the *secondary use of*

data beyond the original purpose for collecting the data. We broadly define community members as *citizens who utilize data to make an impact on the community* (e.g., educational consultants, city planners, social activists, and community foundations), as well as *those who reuse data for community-related decision-making* (e.g., parents looking for educational recommendations for schools for their children). To achieve this goal, we will (1) characterize community members' data reuse practices and strategies and (2) map these practices and strategies to the existing infrastructure focusing on public libraries, in order to better support communities.

Our intended outcomes include (1) providing an understanding of community members' current data workflows to identify barriers and to suggest a model for a community-inclusive data workflow; (2) having a better understanding of the role and the potential of public librarians to support community members' data reuse; and (3) contributing to create a collaborative data reuse culture among the local data infrastructure by helping public libraries design data services that support community members' needs and by providing useful educational resources for librarians to perform such roles. The key outputs will include the following:

- The current workflow of utilizing data for community data reusers with a list of barriers and challenges
- A theoretical model of a community-inclusive data exchange or workflow
- Perceived roles of public librarians in supporting a community's data work
- A list of challenges and barriers of public libraries to offering community data services
- Practical recommendations for public librarians to create a community-inclusive data workflow
- A prototype data service for public library users
- Educational resources for public librarians
- A Public Library Association (PLA) workshop for librarians

Intended Audience

The outcomes expected from the project will serve three primary audiences.

1. ***Community members (citizens)*** who are interested in reusing data for the development of the community and for making decisions relevant to their lives. Our project intends to help citizens connect with existing data resources and have the ability and skills to utilize data by providing education and services through libraries. In particular, providing underserved communities with the opportunity to access and understand data that affect their lives is imperative given the long history of manipulation of these communities by researchers. Several cases (e.g., the Tuskegee syphilis experiment, the exploitation of Henrietta Lacks, and the lead poisoning water crisis in Flint, Michigan) illustrate this need (Leduff, 2016; Skloot, 2011).
2. ***Local data infrastructure and organizations*** (including CRCs) that provide data services and help diverse types of data reusers, including community members. Our project intends to help build a collaborative and community-inclusive culture of data reuse among different stakeholders. Each data infrastructure and organization can complement others by sharing information or resources and serving the needs of different user communities.
3. ***Public librarians*** who want to build the capacity to meet the data needs of community members. Public librarians will benefit from this project by learning how to provide the services and resources necessary to empower library users by increasing their skills (e.g., data literacy skills) and interests. Contributing to an informed and engaged citizenry is a longstanding goal of U.S. public libraries. Given the increasing prevalence of data-driven decision- and policymaking (Kitchin, 2014b; Shirley, 2016), if public librarians do not build the capacity to reuse data, the librarians will face considerable obstacles to meeting this goal.

Research Design

To address the research questions proposed above and achieve the project's goals, multiple research methods will be employed with different activities. See **Supporting Document 1. List of Project Activities** that lists the specific research questions, the methods associated with the intended outcomes, and project outputs.

Phase 1. Study of community data reuse practice – Interview, survey, and theoretical model development (Months 1–14: Oct. 2017–Nov. 2018)

1.1. Interview with community data reusers: IRB preparation and interview protocol development (month 1), interview data collection (months 2–4), interview data analysis (months 5–7)

To understand community members' data practices and reuse workflow, such as the processes, strategies, and challenges (RQ1), we will interview community data reusers. To identify community data reusers, we partnered with 10 CRCs in different U.S. cities and will use the CRCs' networks. As the preliminary study suggested, regional CRCs help connect community members with data sources, partnerships, and consultations. From our initial search, we identified about 50 CRCs, which exist as nonprofit organizations or are affiliated with universities. We also reviewed the 42 partner organizations in 30 cities under the National Neighborhood Indicators Partnership (NNIP, http://www.neighborhoodindicators.org/), which is a collaborative effort by the Urban Institute (http://www.urban.org/) and local partners to further the development and use of neighborhood information systems in local policymaking and community building. We carefully reviewed each CRC's website regarding their missions, services offered, and past projects and contacted them for a potential partnership-building process. After assessing the fit for our project, we chose 10 CRCs in small, medium, and large cities (e.g., Grand Rapids, MI; San Antonio, TX; and Boston, MA) as our partners. Each partner CRC agreed to provide a list of community data reusers among their network and service users. We will interview 2–5 data reusers from each center, about 30 in total. Since each CRC has distinctive community and service users, we did not ask for a certain type of data reuser to be interviewed. However, we asked each CRC to recommend community data reusers who represent diverse community members or data reusers. To complement our understanding of community members' data practices, we will also interview key staff of CRCs, as well as regional data infrastructures (e.g., data repositories and other institutions) that may be referred to by the community data reusers, and explore their role in supporting community members' data work.

The interviews will be semi-structured. The interviewer will rely on pre-developed questions to collect responses for the identified research questions, but the interviewer will further investigate any topic of interest by probing. The interview data will be audio-recorded and then fully transcribed by professionals. We will use NVivo, a qualitative data analysis tool, to code the transcripts and then analyze them. Because the interview study aims to "understand and describe as accurately as possible the phenomenon (data practices)" (Groenewald, 2004, p. 5), we will follow a phenomenological and interpretive approach for analysis, inductively analyze the data and develop codes, and see whether a theme emerges (Smith, Flowers, & Larkin, 2009).

1.2. Survey with community members: Survey design and IRB (months 8–9), survey data collection (months 10–11), survey data analysis (months 12–13)

Once we complete the interviews with community data reusers, we will design and conduct a survey with the broader general population of the communities. Our survey instrument will be devised based on the findings from our interviews. The goal of the survey is to fully answer the research question by addressing community members' needs and experiences with data where CRCs or other types of data infrastructure may not exist or are less accessible. This will be done by including communities with diverse geographic and socio-economic characteristics and by including the perspectives of community members, who are not aware that they are reusing data or who are interested in using data but have not done so for various reasons.

For reasons of convenience, our survey will be administered in diverse geographic regions of Indiana. We examined the list of public libraries from the Indiana Public Library Directory (http://www.in.gov/library/files/cityindex13.pdf) and selected 9 public libraries based on geographic (rural, suburban, and urban) and ethnic diversity based on Census data. Our selections include metropolitan areas (Fort Wayne, Gary, and Indianapolis), as well as small towns near the Indianapolis region (Hammond, Carmel, Greenfield, Vevay, Warren, and English) to capture the suburban and rural demographic. We will work actively with the Indiana Statistical Consulting Center (ISCC) to determine the appropriate sample size based on each city's population and the number of library users, and the statistical power to ensure that our results are generalizable. Through connections with the school's alumni working in public libraries throughout the state of Indiana, an online and on-site survey will be administered through public library websites and reference desks. To increase the survey response rate, a small incentive in the form of a $30 Amazon.com e-gift card will be awarded to 10 randomly selected participants upon completion of the survey.

For the online survey, we will use Qualtrics or SurveyMonkey, an online survey application under the Indiana University account, in order to collect the data. The survey will be linked to each public library's website. For the paper survey, we will share the survey instruments with library directors or reference librarians

by email or by mail for printing depending on each library's preferences, so that they can place the instrument at the reference desk. Once the survey is closed, we will collect the paper surveys by mail or in person. The project assistant will code the survey results in CSV format, which will then be merged with the online survey results that will be exported from Qualtrics or SurveyMonkey in CSV, and then clean and recode the data to import them into SPSS or Stata for statistical analysis. As the goal of this survey is to describe the practices and perceptions of community members, data analysis will be limited to descriptive measures, and the data will be analyzed using univariate and bivariate descriptive statistics (e.g., frequency distribution).

1.3. Theoretical model development with practical recommendations (month 14)

Using the findings from interviews and surveys, the project team will create a theoretical model of, and practical recommendations for, a community-inclusive data exchange and workflows. The purpose of developing the model is to suggest a method for improving community members' data workflow by presenting different stakeholders and local data infrastructures that can or should be involved in the community's data work. This model will contribute to broadening understanding of the data life cycle, which mostly reflects the data practices of the scientific community. The model will also suggest how the organizations can work together or complement each other to support the community's data utilization. The theoretical model and practical recommendations will be sent to our advisory board for their initial review and feedback, and they will be updated at the end of Project Phase 2, with the completion of an interview study with public librarians.

Phase 2. Study with public librarians: Environmental scan, interview with librarians, and partnership building (Months 15–25: Dec. 2018–Oct. 2019)

2.1. Environmental scan: Environmental scan of library work (months 15–16)

We will conduct an environmental scan to identify public libraries and library services from around the country that are currently engaged in data initiatives. The purpose of this environmental scan is to identify librarians working with data for their community's use in order to invite those librarians to participate in our interview study in Phase 2. Several strategies will be employed: First, we will work with individual state libraries' public library services directors. Public libraries are required to submit strategic plans, as well as technology plans, to qualify for Library Services and Technology Act (LSTA) funds from the state library. These plans provide a centralized location for obtaining information regarding services offered at public libraries across the nation. Second, we will conduct an online search to find relevant news articles and will review relevant conference programs. Third, as our advisory board includes representatives of public libraries who are involved in data initiatives and who have a broad professional network, they will also help in recruitment by suggesting potential participants and referring to a professional community for the project team to reach out.

In addition to librarians currently working on data initiatives, we will identify business and community outreach librarians for interviews. These individuals routinely work with data and/or with community issues throughout the course of their work. We will identify librarians who work at library systems with $5 million or more in annual expenditures. Systems of this size typically employ librarians with these specializations, and these types of librarians routinely work data and/or with community issues throughout the course of their work.

2.2. Interview with librarians: IRB preparation and interview protocol development (months 17–18), interview data collection (months 19–21), and interview data analysis (months 22–24)

We will conduct interviews with about 15 librarians identified through the environmental scan. The purpose of the interviews with librarians is to learn about their current and/or perceived roles to support their community's data needs, current efforts and services, challenges and opportunities, knowledge and skills librarians have or need, any resources needed and their opinions about effective data services deployment (RQ2). The interviews will also be in a semi-structured format, following pre-developed questions while allowing the interviewers to further explore areas of interest that appear during the conversation. We will again use NVivo to inductively analyze the collected qualitative data using an open coding method, from the phenomenological and interpretive approach (Smith et al., 2009). We will also review the services mentioned by the librarians (if any pertinent online information is available) in order to better understand the types and nature of the services.

2.3. Model revision and recommendations update (months 24–25)

From the qualitative analysis of the interview data and the examination of services introduced during the interviews, we will update the theoretical model and the practical recommendations developed in Phase 1 to reflect the roles of public libraries. The updated models and recommendations will be reviewed again by our advisory board for their feedback. The outcomes will also be distributed online through the listservs of library associations in order to seek input from members of the Public Library Association (PLA) and the American Library Association (ALA) at the start of Phase 3 (see Phase 3 for details).

2.4. Phase 3 preparations (months 15–25)

- **PLA Workshop:** Proposal application for PLA 2020 (months 19–20)
- **Partnership building:** Recruitment of participating libraries (months 15–25)

During Phase 2, we will also start recruiting the participating libraries for Phase 3, library service design. The environmental scan will help us to identify partner libraries that have implemented efforts to support community data needs or that are interested in building data services for their community. We will also attend the ALA and other conferences to advertise our project and recruit partners.

Phase 3. Public library service prototype design and workshop (Months 26–36: Nov. 2019–Sept. 2020)

3.1. Phase 2 output distribution for community review (months 26–27)

In Phase 3, we will utilize evidence from our findings to improve or develop library services. We will first distribute our Phase 2 outputs for community review. For this community review, we will distribute the outputs online through the listservs of library associations and relevant committees or sections in order to seek input from members of the Public Library Association (PLA) and the American Library Association (ALA) at the start of Phase 3. PLA members will be generally targeted, along with the PLA's Technology Committee, as well as two sections of the ALA's Reference and User Services Association (RUSA): the Business References and Services Section and the Reference Services Section.

3.2. Prototype service design: Development of prototype service design (months 28–33)

We will work with the partner libraries recruited during Phase 2 to design a prototype user service. We cannot specify what types of service we will be designing; our goal is to let the data guide us with the project's progress. Prototype services for public library users would evolve alongside the theoretical model. However, at this point in time, given existing research, we know that data literacy skills, which are fundamental to any type of data work and reuse, as well as data analysis/visualization instruction and resources, will be necessary parts of these services (Herzog, 2015; Macy & Coates, 2016; Evergreen, 2017).

3.3. Library educational resources: Development of educational resources for librarians (months 28–33)

We will also compile and develop an educational resource for librarians throughout the project and provide access through a project wiki. The wiki will provide access to workshop materials, descriptions of roles and competencies, a bibliography, and related resources.

3.4. Library workshop: Workshop preparation, advertisement program design and post-workshop evaluation question design (months 28–30), evaluation survey with workshop participants (month 31), and survey data analysis (month 32)

Dissemination of the project outputs is important to increase the chance of having an impact on the library community. We will present our findings at the PLA and other relevant conferences to widely disseminate the findings. We will also run a half-day workshop at the PLA preconference. The purpose of the workshop is to bring librarians into broader conversations about their community data work.

This workshop is designed to provide librarians with an overview of community members' data work and the potential role of public libraries in serving community members' data needs. Participating librarians will explore opportunities for libraries to actively engage with data and community members' data work in their organizational context, and transfer their knowledge and experience of serving their community's information needs into a new domain, the community's data needs. In this workshop, there will be brief presentations from librarians who have experienced with data initiatives (one of our advisory board members), as well as librarians who will participate in the prototype data service design (two librarians from our library partners). Participants will engage in hands-on activities, guided discussions, and a brief Q&A with speakers while networking with

other participants with the same interests to exchange ideas and gain insights. We aim to have about 30 participants in this workshop, and participants will receive pre-workshop materials (e.g., background and bibliography), slides, exercises, and the schedule before the workshop. The workshop will be divided into two sessions. The first session (80 min) will focus on opportunities for public libraries to support the community through data (conceptual grounding). The first session will be structured as follows:

- Introduction: Motivation for the workshop and the project, cases that demonstrate the impact of data on communities (project team)
- Presentation: Project findings from Phases 1 and 2 (project team)
- Speaker presentation: Experiences and perspectives of a library that was previously involved in the data initiative for the community, followed by a Q&A (advisory board member from the project)
- Group discussion on the potential role of libraries to serve community members' data needs and opportunities for librarians to be involved in data work (all)

The second session (80 min) will delve into the role of public libraries in building the community data infrastructure and service design and provide practical implications for designing services with hands-on activities. The second session will be structured as follows:

- Brief presentations from our partner libraries that will develop prototype services in this project with a Q&A (two librarians from our partner libraries)
- Guided discussion on barriers and challenges to designing and implementing the services; problem-solving activities to identify common strategies, approaches, and necessary support to build a data service in libraries (all)
- Group exercise on library service design following the example prototype services developed in this project

At the end of each session, we will administer a post-workshop evaluation to the workshop participants to collect feedback from the participants. All workshop materials will be freely available through the project website. The project team will also document the discussions and encourage collaborative note taking, which, in addition to the materials used in the preparation of the workshop, will be freely available.

3.5. Final report and output dissemination (months 34–36)

The project team will write a report to submit to IMLS and prepare to disseminate the project findings and output of the project (e.g., project information, our preliminary work, as well as final work) through multiple channels, although outcome dissemination will take place throughout the project periods.

Diversity Plan

The composition of our project team (PIs and advisory board members) is diverse in terms of gender and cultural/ethnic background. We expect that having a diverse project team will be beneficial in bringing a variety of perspectives to bear on all of the phases of this project, and we will actively recruit graduate research assistants who can enhance diversity. We will also address the need for diverse viewpoints throughout the project in order to be as inclusive as possible in our approach. For the interview study with community members, we will stress the importance of having diverse perspectives when communicating with the CRCs regarding community data reusers' recruitment. For the survey study with community members, we will employ several strategies to include the perspectives of a geographically and socio-economically diverse community. Our selections of 9 cities where the survey will be conducted is based on geographic and ethnic diversity, and they differ in size and population to capture community members' perspectives from metropolitan, suburban and rural areas. In Phase 2, we will also recruit librarians who are involved in data reuse initiatives and in the provision of business reference services or community outreach services from a variety of geographic regions that reflect ethnic, racial, and economic diversity.

Project Evaluation

The project evaluation will consist of two parts: an evaluation of the research components and an evaluation of the project progress and success. For various research components, the project team will examine the validity

and reliability of the data collected from the interviews and surveys. Inter-coder reliability will be checked for open coding of the interview data. To evaluate the project's progress and outcomes, we will adopt a participatory evaluation framework and invite our advisory board and our partners to evaluate the project during each phase. First, our advisory board (a team of experts from public libraries, community research centers, academic libraries, and community outreach) will participate in the ongoing evaluation (e.g., reviewing the development of the instrument, the implementation processes, and the team's collaborative work with partners). Each year, the advisory board will be invited to Indianapolis for a half-day annual evaluation. After project team's brief report on the project's progress, the advisory board will be asked to fill out a paper evaluation form regarding the project's process, timeline, and outputs, which will have a mix of Likert-scale and open-ended questions. Then, the project team and the advisory board will have a further discussion regarding the project's outputs and impact, including how our findings reflect or contrast with their own experiences; how actions or plans to connect users to data would work in their organizations, and whether the resources and services developed are flexible enough to meet diverse institutional or community needs. Second, we will also collect an online evaluation, in the form of Likert-scale and open-ended questions, from our CRC and library partners, regarding their general experiences working with our project, communication and collaboration, usefulness of the outputs, the perceived impact, and outcomes. Similarly, we will collect post-workshop evaluations from our workshop participants regarding the usefulness of the workshop, content, and impact. Finally, as we will share the project information and products on a project website, we will also keep monitoring the visit and download numbers of the project materials on the website and count the number of participants in the workshop and conference sessions. Additionally, a long-term indicator of success for the research components will be measured from the number of citations and the circulation of publications as a way of quantifying the success of the project.

Project Resources: Personnel, Time, Budget

Project Team

Ayoung Yoon, Ph.D. (PI) is an assistant professor in DLIS at IUPUI and is a RDA/US data share fellow. Her dissertation, *Data Reuse and Users' Trust Judgments*, received the Eugene Garfield Doctoral Dissertation Award. She has been actively involving in data curation and data sharing/reuse research, and her ultimate research goal is to construct effective community inclusive cyber infrastructure for data sharing and reuse through proper curation. She published a number of articles in top rated journals, such as *Journal of the Association of Information Science and Technology* (JASIST), *College & Research Libraries* (C&RL), and *International Journal of Digital Curation* (IJDC). Her previous work has been funded by Indiana University, University of North Carolina at Chapel Hill, and the Alfred P. Sloan Foundation.

Andrea Copeland, Ph.D. (Co-PI) is an associate professor in the DLIS at IUPUI. For over a decade, she has researched the relationship between communities and public libraries. Current emphasis is on connecting the cultural outputs of individuals and community groups to a sustainable preservation infrastructure. She is the co-editor of the volume, *Participatory Heritage* published by Facet in 2017.

Advisory Board Members (**Support Letters** are attached in **Supporting Document 4**)

Nate Hill is the Executive Director of the Metropolitan New York Library Council (METRO), a multi-type library consortium serving over 250 organizations in New York City and Westchester County. He had involved in the Open Data initiative when he served as Deputy Director of the Chattanooga Public Library.

David Leonard is the President at Boston Public Library. He has been involved in Boston Public Library's recent initiative to develop open data collection for the City of Boston.

Rebecca Stavick is the Executive Director of Omaha's first digital library, named Do Space. In 2012, Stavick co-founded Open Nebraska, a local organization focused on civic application development, open data advocacy, and technology education.

Ellen Cutter is the former Director at Community Research Institute at Fort Wayne and currently the Director of Strategy and Research at Greater Fort Wayne Inc. She is passionate about telling stories with data, tracking a multitude of economic data points to better understand the opportunities for Allen County's future.

Megan Sapp Nelson is a Professor of Library Sciences at Purdue University Libraries. Her research investigates the skills necessary to efficiently and effectively manage data and the pedagogy that can most effectively teach those skills. She was a Co-PI of the Data Information Literacy project funded by IMLS.
John Helling is the Director of Public Services in the Indianapolis Public Library. His roles include using data to help calculate the community's return on investment in public libraries.

CRC Partners (**Letters of Commitment** are attached in **Supporting Document 3**)

1. Community Research Partners (CRP), Non-Profit, Columbus OH, Contact person: Lynnette Cook, Ph.D. Executive Director
2. Community Engagement & Research & Community Assistance Research (CARE), University of Texas Southwestern Medical Center, Dallas TX, Contact person: Sandi Pruit, Ph.D.
3. The Polis Center, IUPUI, Indianapolis IN. Contact person: Sharon Kandris, Associate Director
4. Community Research Center, Tufts University, Boston MA, Contact person: Doug Brugge, Director
5. Data You Can Use, Non-Profit, Milwaukee WI. Contact person: Kathleen Pritchard, Ph.D. President
6. Center for Community and Business Research (CCBR), University of Texas San Antonio, San Antonio TX, Contact person: Thomas Tunstall, Ph.D. Senior Director
7. Southeast Community Research Center, Non-Profit, Atlanta GA, Contact person: Douglas Taylor, Ph.D. Director
8. Community Research Institute, Dorothy A. Johnson Center for Philanthropy, Grand Rapids, MI. Contact person: Jodi Petersen, Ph.D. Director
9. DataHaven, Non-Profit, New Haven CT. Contact person: Mark Abraham, Executive Director
10. DataSpark RI, Non-Profit, Providence RI. Contact person: Kim Pierson, Director

-
-
-

-
-

Communication Plan

There are four main thrusts to disseminating information about this project: 1) communication with community members who are interested in using data, public libraries, scientific researchers and the wider public regarding the importance of data for community development and citizens' lives as well as the current workflow of community members' data work; 2) communication with the library professions and the wider public about the library prototype services; 3) communication about and recruitment for partner libraries at Phase 2; and 4) communication with the library professions concerning the PLA workshop. The project team will develop numerous PR documents and communications in disseminating information for each of the above strands. The project team will use several approaches to project information and disseminate outcomes including the following: conference papers and presentations, journal articles, a project website, our partner organizations as well as collaborators, project teams' professional networks and affiliations with professional associations, the DLIS alumni network and through extensive event (workshop) and outcome announcements on a wide range of listservs for library audiences.

Sustainability Plan

The impact of this research beyond the project people will depend on how successful we are at building capacity throughout the project. First, it will be important to educate all parties about the significance of the

others in the data workflow landscape. CRCs and public libraries contribute different and valuable resources, perspectives, and skills to that landscape. The more each constituent is aware of the others' contributions the better communities will be served be both. Through the education resources we produce, awareness will be increased. Secondly, by working with librarians to co-create professional development resources to facilitate the provision data reuse services in public libraries, the adoption of these resources is more likely and with that their future evolution more likely. Upon the completion of this project, the project team will apply for the full National Leadership Grants for Libraries, National Digital Platform category to fully develop the user services with our partner libraries, which can be a model for other libraries that want to develop a service.

NATIONAL IMPACT

This project meets the IMLS agency-level goals **by supporting libraries to be strong community anchors that enhance citizens' engagement with data and create opportunities for citizens to learn about and work with data for the development of the community.** Belonging to the category of **Community Anchors**, our project will advance the role of libraries in connecting citizens with existing data sources and infrastructure while promoting awareness of the potential for data to be used for community development and for providing necessary services (e.g., data literacy education) to the public.

Specifically, the project will have major impacts in several areas. In terms of research, this project will contribute to broadening the scope of data reuse research because overwhelmingly data reuse research focuses on scientific and academic contexts. As the practice of data reuse is expanding with many open data initiatives in public sectors, it is timely to investigate data reuse practices outside the academic context, particularly from the community and citizen perspectives. Our research will provide comparative insights for understanding data reuse practices in different contexts and will help to set the next research agenda to bridge the connection between scientific and local community data reuse.

Second, the project will help build a collaborative, community-inclusive culture of data reuse among different stakeholders involving public libraries. Our research will show how public libraries and other types of data organizations can complement each other to serve the community and how the local data infrastructure and organizations can work together by bringing their expertise to promote knowledge and resource sharing within the community. Because CRCs, as regional non-profit organizations or academic institutions, play a significant role in connecting community members and organizations to the scientific community, there is great potential for CRCs to collaborate with public libraries to promote the use of data at the local level. Public libraries routinely partner with other community agencies with complementary services to the benefit of their respective communities. This will serve the public and/or different populations that may not always be served by CRCs. The types of data and services offered by public libraries and CRCs vary in large part based on the types of users they typically serve.

Last, our work will offer insights into how current data services can be adapted or how data services can be newly developed to meet the needs of community reusers. In particular, our project will raise awareness of the role of public librarians in community data reuse through the inclusion of their expertise, as well as that of community stakeholders in the design or extension of library data reuse services. Although understanding community members' data needs and practices is the first step for developing a library's data services, it is equally important to provide education or training for public librarians so that they can be equipped with necessary skills and knowledge and further develop their expertise in new areas. With these skills and knowledge, public librarians will be able to provide educational programs to the communities they serve for utilizing data (e.g., data literacy education, which is fundamental to successful data reuse as data literacy skills will provide citizens the basis for critically examining and using/reusing data). Several studies (e.g., Kitchin, 2014; Borgman, 2015) have addressed the social, political, and ethical consequences of the data revolution; however, as yet, there are no resources for citizens to understand these consequences in so far as they relate to community members' everyday lives. Our project will address the need to lift barriers that prevent community members from understanding the analyses going on around them that could affect their lives by developing educational resources for librarians, developing prototype user services for community members, and educating community members (citizens) for data reuse through library services.

Schedule of Completion: Year 1

Phase	Activities		2017 Oct	Nov	Dec	2018 Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep
Phase 1	Interview study: Community members	IRB & Interview protocol development												
		Data collection												
		Data analysis												
	Survey study	Survey design & IRB												
		Data collection												
		Data analysis												

Schedule of Completion: Year 2

Phase	Activities		2018 Oct	Nov	Dec	2019 Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep
Phase 1	Survey study	Survey design & IRB												
		Data collection												
		Data analysis												
	Theoretical model development													
Phase 2	Interview study: Librarians	Environmental scan												
		IRB & Interview protocol development												
		Data collection												
		Data analysis												
	Theoretical model revision													
	Phase 3 preparation	Workshop proposal												
		Library partnership building												

Schedule of Completion: Year 3

Phase	Activities		2019 Oct	Nov	Dec	2020 Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep
Phase 2	Interview study: Librarians	Environmental scan												
		IRB & Interview protocol development												
		Data collection												
		Data analysis												
	Theoretical model revision													
	Phase 3 preparation	Workshop proposal												
		Library partnership building												
Phase 3	Community review of outcome													
	Prototype service design													
	Library educational resource development													
	Workshop	Preparation												
		PLA Workshop												
		Evaluation & analysis												
	Final report													

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions
You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

PART I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

This project will produce datasets coming from interviews and surveys. The datasets will be in the format of text files and spreadsheets. The PIs will hold the copyright of the datasets. The datasets can be reused for non-commercial purposes. The following Creative Commons license will be assigned: CC BY-NC. The dataset will be deposited in IUPUI ScholarWorks, which is IUPUI Institutional Repository (https://scholarworks.iupui.edu/).

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

The copyright of the datasets will be held by the PIs. The PIs will allow the sharing of the datasets for non-commercial use. Creative Commons will be used to describe the conditions of access and use. The following license terms will be applied:

- Share - copy and redistribute the materials in any medium and format;
- Adapt - remix, transform, and build upon the material;
- Attribution - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use;
- NonCommercial - You may not use the material for commercial purposes (http://creativecommons.org/licenses/by-nc/4.0/)

The licensing information will be notified when any users attempt to download the dataset from the repository.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

N/A

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

N/A

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.
N/A

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

N/A

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

N/A

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

N/A

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

N/A

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

N/A

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

N/A

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

N/A

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

N/A

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

N/A

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

N/A

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

N/A

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

N/A

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

N/A

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

N/A

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

N/A

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

N/A

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

N/A

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: N/A

URL:

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

This project will collect empirical research data through interviews and surveys to achieve the research goals. Interviews with community data reusers will be conducted to understand community members' data needs and practices, followed by the interviews with key staff at local data infrastructure including community research centers (Oct 2017-Apr 2018). To fully uncover the research question by addressing community members' needs and experiences with data where CRCs or other types of data infrastructure may not exist or are inaccessible, a survey will be administered at 9 public libraries in Indiana, which are selected based on geographic (rural, suburban, and urban) and ethnic diversity (May 2018-Oct 2018). Another interviews will be conducted with public librarians to learn about their current and perceived roles to support their community's data needs and work, current efforts and services, challenges and opportunities, knowledge and skills librarians have or need, any resources in need and their opinions about effective data services deployment (Feb 2019-Sep 2019). The collected data will be analyzed quantitatively and qualitatively to accomplish the project goals.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

The data collected from interviews and surveys will require approval from the Institutional Review Board (IRB) at Indiana University. We will submit IRB documents for interviews with community members in Oct 2017 (Month 1), for surveys with community members in Jun 2018 (Month 9), and interviews with public librarians in Feb 2019 (Month 17). If all will be treated as exempt from IRB oversight, it will be approved within 2 weeks. Both PI and Co-PI have had multiple successfully approved IRB applications.

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

The interviews and surveys will collect demographic information. We will not disclose and de-identify any personally identifiable information (e.g., name of person and organization) from all interview data. All subjects will be assigned a randomly generated ID number. For survey data, we will also use identification numbers to ensure the anonymity of responses and all responses will be aggregated in the analysis. Before releasing the collected data to the public, we will remove any information that might include personal information, and ID numbers will be used for each response. All these processes will be approved by the Indiana University Office of Human Subjects prior to data collection.

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

We will collect informed consent forms for interviews and surveys. The consent forms will include the description of the project, participation process, confidentiality, compensation, and other information. In addition, the consent form will inform the participants our data deposit and sharing plan. Only those who agree to the consent form will be able to participant in the study.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

For the narrative data from interviews, we will create audio files that record dialogues in mp3 format. We will hire professionals who will transcribe the audio into text files (e.g., .docx, .txt).

For collecting survey data, we will use the Qualtrics or Surveymonkey under IUPUI Department of Library and Information Science account, which provides an online survey template. The collected survey responses will be downloaded in the format of .CSV, and will be analyzed using the SPSS, a statistical analysis software program (https://www.ibm.com/us-en/marketplace/spss-statistics).

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

Interview protocols, survey questionnaires and coding book for surveys will be created along with our datasets. Those documents will be stored and managed together with the associated data in the IUPUI ScholarWorks. Any user who is interested in the dataset will be able to access those additional documents along with the datasets.

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

After the completion of the project, our interviews and survey data will be archived, managed, and disseminated through the IUPUI ScholarWorks. Relevant metadata will be added to make them easily searchable on the web. In addition, we will include the information about dataset access in our publications and will also share the links to the datasets on our project website.

A.8 Identify where you will deposit the dataset(s):

Name of repository: IUPUI ScholarWorks

URL: https://scholarworks.iupui.edu/

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

Data management plan will be reviewed every time we create and store data. Specifically, we expect that we will review this plan at least three times when we complete collection of interview and survey data. We will do a final review of our data management plan upon completion of this project.