Title: Curating Data to Enhance Public Library Effectiveness

The Public Libraries Survey (PLS) is an annual survey of U.S. public libraries. Policymakers and practitioners depend on PLS data to allocate funding and strategically manage libraries. Academics rely on PLS data to conduct original research about public libraries' effectiveness. Data in the PLS come from over 17,000 outlets, and it represents a "gold standard" for national information about public libraries (IMLS, n.d.).

While the PLS is an invaluable resource for the public library community, other organizations collect data that may extend the reach and significance of the PLS. The Association of Bookmobile and Outreach Services (ABOS) Bookmobile and Outreach Information Repository (BOIR) contains unique data about library outreach services unavailable in the PLS. The Public Library Association's (PLA) Project Outcome (PO) initiative tracks the impact of programming in the communities that libraries serve. As noted by Measures that Matter (MtM), an initiative funded by the Institute of Museum and Library Studies (IMLS), "**there is no overarching national plan for the collection, storage, use, and dissemination of public library data and service outcomes**" (Smith et al., 2017). Organizations like ABOS and PLA collect multiple information sources and formats, making fragmented datasets inaccessible to the public. Nationally, making changes to the PLS is time-consuming (Linda Hofschire, Director of the Colorado State Library's Library Research Service, personal communication, November 2020), which makes federal approaches to collecting and disseminating information about library service outcomes difficult.

In response, the Inter-university Consortium for Political and Social Research (ICPSR) requests **$249,682** for an **Implementation Grant** from the IMLS. We propose aggregating, curating, then archiving a one-of-a-kind dataset that documents the impact of library programming and outreach on communities, simultaneously prototyping a collaborative model to help address the lack of a national strategy noted by MtM. Combining ABOS and PO data then housing them in a single location will enable library administrators and researchers to examine the results of library outreach and programming decisions on a more granular basis, contributing to the completion of MtM's Action Plan goals (Wiggin et al., 2018).

ICPSR has partnered with three organizations to complete this project: ABOS, PLA, and the School of Information Science & Learning Technologies (SISLT) at the University of Missouri. ABOS and PLA will provide data to ICPSR, which will process, enhance, and archive it for subsequent reuse. SISLT will oversee the creation of a data module that augments the PLS by linking it with ABOS and PLA data at the library administrative unit level, identifying new variables to compute, adding new data points from other datasets, and documenting transformations we carry out. Researchers may use our data module to merge curated data with the PLS to maximize analytical value. ICPSR, PLA, and SISLT will advertise project findings, promote data dissemination, and train graduate students to collaborate across professional networks to use and leverage data.

Project Justification: This project addresses National Leadership Grants for Libraries, **Goal 5**, **Objective 5.1,** by strengthening libraries' capacity to work collaboratively to benefit the communities they serve. Reviewing individual data elements collected by state libraries, MtM found that "on average, states add 180 additional data elements" beyond those collected by IMLS for the PLS (Blankstein & Wolff-Eisenberg, 2019, p. 2). These added data elements suggest public library stakeholders need more data than the PLS provides. Organizations like ABOS and PLA collect data, demonstrating public libraries need additional data to expand on existing PLS data.

Unlike other countries, the U.S. statistical system is decentralized, with no one agency responsible for national data work (Fecso, 2012). The vast majority of public library funding originates at the state and local levels (Million & Bossaller, 2020), and the 10th Amendment of the Constitution delegates unenumerated powers to state governments. Consequently, U.S. public libraries are local entities often responsible for gathering their own statistical information, sometimes with support from professional associations (e.g., the American Library Association or PLA). However, these

datasets are often fragmented, and IMLS is one of the few entities with the resources and authority to aggregate and normalize national library statistics.

The PLS is a voluntary survey that provides data about public libraries in the U.S. The PLS is IMLS' default mechanism for collecting national statistics. The survey has been conducted annually since 1988, and it is administered by data coordinators appointed by the heads of individual state libraries. Data are collected using a web-based reporting system (IMLS, n.d.). **However, data releases lag by several years from data collection, and our discussions with library stakeholders indicate that adding items to the PLS is not always feasible.** This stems from the *Paperwork Reduction Act*, which requires the U.S. Office of Management and Budget to review individual survey changes and the difficulty of bringing 50 states together to determine variables for the PLS to collect.

The PLS provides an excellent system for collecting and normalizing data, but public libraries do not have a national strategy for collecting, curating, validating, and sharing other data sources. The "data deluge" led to a dramatic increase in the amount of data available to library stakeholders (Hey & Trefethen, 2003), with information collected by state libraries, ABOS, and PLA each an example. Combining ABOS and BOIR data in one place and augmenting the PLS with this data would provide valuable information to libraries and create a prototype for national library data governance.

Table 1. *Key Acronyms in this Proposal*

Acronym	Name
ABOS	Association of Bookmobile and Outreach Services
BOIR	Bookmobile and Outreach Information Repository, an initiative of the Association of Bookmobile and Outreach Services
DDI	Data Documentation Initiative
DPN	Digital Preservation Network
ICPSR	Inter-university Consortium for Political and Social Research at the University of Michigan
IMLS	Institute of Museum and Library Services
ISO	International Organization for Standardization
OAIS	Open Archival Information System
PLA	Public Library Association
PLS	Public Libraries Survey, the primary data collection survey conducted by the Institute of Museum and Library Services
PO	Project Outcome, an initiative of the Public Library Association
MtM	Measures that Matter
SISLT	School of Information Science & Learning Technologies at the University of Missouri

<u>Connection to Theory, Scholarship, and Practice</u>: This project supports a need for data-driven librarianship indicated by 32% of respondents to a survey of the public library field (PLA, 2021). This survey found commonalities in community priorities, demonstrating the need for libraries to compare and contrast outcomes in areas such as assisting job seekers based on different inputs and contexts. Massis (2016) points to data-driven decision-making as a way in which libraries can demonstrate "radical transparency" (p. 131) and share the results of their decisions. This same philosophy of transparent decision-making motivated PLA's Project Outcome to assist libraries in measuring the benefits their services provide to communities.

This project also expands on work done to integrate library data with other data sources. Prior research at the national level has not focused on integrating public library outcome data with the PLS. In contrast, other projects have focused on single library systems and excluded rural and small libraries. GEOLIB, an IMLS-funded project from the 1990s, connected Census data with library use data to demonstrate how user demographics influence service uptake. This project demonstrated the potential for using geographic information systems, Census demographics, and librarians'

local knowledge to improve services. Commercial services such as OrangeBoy have taken up this mantle and are used by public libraries to understand patron demographics; however, they do not use library-generated data. Rosichan (2020) demonstrates a practical potential for combining datasets locally. The Nashville, Tennessee public library uses demographics and local library data for evaluation purposes. Ostler, Norlander, and Webber (2021) demonstrate the potential for data integration to enhance branch library services at the Seattle Public Library. They note "most public libraries simply do not have the capacity to meaningfully engage with datasets" (p. 374), suggesting a need for assistance from outside groups. These projects demonstrate the value of connecting and integrating library data. Still, previous efforts have failed because a national, collaborative model for dataset aggregation and dissemination to complement the PLS does not exist.

How We Address the Problem: We address the lack of a national plan by libraries to handle service and outcome data by prototyping a model to aggregate, curate, and archive ABOS and PLA data then connect it with the PLS. The PLS only collects **input** and **output** data (e.g., collections budgets, materials circulated), and practically no data regarding outreach efforts other than bookmobile counts. ABOS' BOIR tool provides unique data about **outreach** (e.g., outreach programs offered, program audiences, bookmobile capacity), capturing a more comprehensive depiction of services libraries provide to their communities. ABOS is uniquely positioned to collect this information because it is an association dedicated to promoting "relevant and responsive services to individuals and groups who face barriers to library access" (ABOS, n.d., para. 1). PLA's PO data contains 357,828 observations from 2,075 public library patron ratings of civic engagement, early childhood literacy, economic development, education, and public health programming (PLA, 2017). PLA (n.d.) is the largest national association "dedicated to supporting the unique and evolving needs of public library professionals" (para. 1).

ABOS and PLA data are **ideal for prototyping a model** to aggregate and evaluate public library service data. These data are: 1) not publicly available, 2) have value, and 3) already exist or are being collected. Unlike unique data collected by state libraries, these two datasets are national in scope. However, BOIR and PO data also contain information that describes individual libraries and library districts, making them an optimal choice for contributing to MtM's Action Plan goals (Wiggin et al., 2018), given their breadth and depth.

While we could aggregate and curate data from the BOIR and PO only, integrating our work with the PLS expands the utility and value of that data which we achieve by creating a **data module**. Recent research finds data curation is one of the strongest predictors of data reuse (Hemphill et al., 2021). Although we will curate data and save it separately from the PLS, creating a data module will allow researchers to identify relationships between our data and the PLS. A representative example of a data module is the University of Michigan's Food Access Module (FAM) to supplement the Panel Study of Income Dynamics. The FAM "allows researchers to investigate how food establishment availability in the physical environment relates to the health and wellbeing of individuals and families" (Tang & McCracken, 2018, p. 4). For this project, our data module will: 1) link BOIR and PO data to the PLS by defining a shared variable (i.e., library administrative units), 2) document data transformations, 3) identify other variables we calculate or incorporate into our data, and 4) describe how we create our final dataset. **We provide additional information about our data module later in this proposal.**

Linking PO and BOIR data to the PLS will enable a deeper analysis of library services. For instance, the PLS measures the number of children's programs that library districts offer and the number of attendees for children's programs (IMLS, 2021). BOIR gathers information about bookmobiles and outreach programming, like figures that describe offsite early literacy programs. PO supplements BOIR data by providing data about patron assessments of programs at libraries. Bringing these data sources together is potentially transformational. For example, a library director could use our work to determine if patrons attend childhood literacy programming in the morning or afternoon, at an outreach site, or in-person at libraries, and if parents feel particular programming types are more effective at helping their children.

Researchers could compare the effectiveness of early childhood literacy programs at the national level based on community socioeconomic variables.
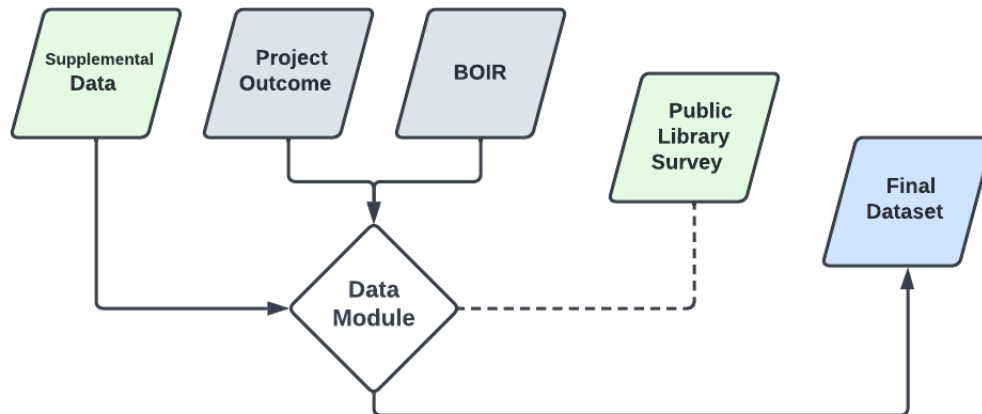

Figure 1. Relationship of Datasets to Data Module and Final Dataset

ICPSR and SISLT's contribution to this project is to provide technical and administrative leadership. ICPSR is the world's largest digital archive of social science data and a winner of the 2019 IMLS Medal for Museum and Library Service. The consortium has 60 years' experience collecting, aggregating, and disseminating sensitive data. ICPSR's Curation Unit employs over 40 data curators with experience processing datasets for the U.S. Department of Justice, the Department of Education, the National Endowment for the Humanities, and many other agencies. ICPSR recently received a $38 million award from the National Science Foundation to build a new generation of research data management tools. **Consequently, ICPSR is positioned to create entirely new processes that leverage already-developed and in-development research data infrastructure**. SISLT will advise ICPSR by providing the subject-matter expertise required to develop our data module and help bring library stakeholders together.

The result of this project will not only be a novel, valuable dataset but a collaborative framework for collecting and disseminating public library data. We will document our lessons learned, solicit regular input from the library community, establish replicable workflows, utilize interoperable data standards to promote data reuse, disseminate our work, and bring together scattered resources to accomplish a shared goal. By leveraging distributed financial, informational, social, and technical capital, this low-cost, high reward project augments the PLS. It also maximizes the value of data that no single party can accomplish alone by fostering collaborative governance and piloting this approach nationally (Ansel & Gash, 2008).

Target Audience: Our primary target audience for this project is public library administrators, state librarians, and researchers. Combining BOIR and PO data with the PLS will allow our target audience to predict which programming and service decisions have significant community impacts. Connecting service outcome data with library revenue and resource data will enable library administrators to explore the variables that make programs effective and examine each in their communities. State librarians can use our data to decide how to invest IMLS funds or deploy statewide services. Researchers will benefit by having a combined dataset that allows a more holistic view of public library practice that includes quantitative and qualitative measures. PLA will undertake training for this target audience by developing webinars and presentations designed to meet the needs of busy working librarians who need focused and just-in-time training.

Our second target audience is library students and educators, whom we will train to use our data and replicate our model for augmenting the PLS with outreach and outcome data. Students will gain skills in working with real datasets,

observe trends and outliers in public library programs and services, and learn how to obtain data for analogous cases in practice. Moreover, through training programs, they will develop networks of colleagues with similar training and concerns. Educators will receive free course materials. In this case, SISLT will lead training efforts by developing learning objects for class use to allow them to explore real-world data and problem-solving.

Beneficiaries: The ultimate beneficiaries of our work will be library patrons whose needs will be better served by programs and services designed to maximize beneficial outcomes. There are 9,057 public libraries in the U.S. (ALA, 2021) that serve the vast majority of the country. Librarians and library administrators have local and contextualized knowledge of their communities' needs and desires. Our project builds on that local knowledge by creating a shared dataset that allows librarians to determine what kinds of interventions work in communities like theirs. While our dataset will be limited, our point-in-time data will enhance libraries' ability to evaluate programs and services for patrons and communities like their own. Furthermore, the public library practitioner community will benefit from our work by learning how to implement a national library data governance model, as will those interested in data librarianship, by consulting our data module.

Project Work Plan: We propose a two-year project with three deliverables:

1. Collect and curate BOIR and PO data
2. Create a data module
3. Create educational materials and conduct outreach to share our data and "lessons learned" with library stakeholders

These three deliverables provide a path forward, and our **Schedule of Completion** contains a project timeline alongside 17 sub-deliverables.

Deliverable 1: Our first deliverable is collecting, curating, and enhancing BOIR and PO data. The BOIR contains unique information about public library outreach departments, services, and vehicles. The PO data includes library programming user evaluations. BOIR data originates from individual library districts across the U.S., voluntarily entered by representatives using a standard web application. PO data comes from libraries across the country that leverage online, validated survey instruments provided by PLA. These data are collected at the end of programming sessions, such as a summer reading program to assess quality. We will ingest both datasets at ICPSR, normalize them, create documentation, and provide them to users via restricted download. At ICPSR, restricted download refers to a process where researchers:

1. Apply to use data
2. Sign an agreement stating they will not publish nor share identifiable information
3. Destroy the data after work is complete

We will restrict access to data to protect the confidentiality of identifiable individuals. Any academic researcher, librarian, member of the press, or policy-researcher who applies will be allowed to access data. We will perpetually host and provide curated data to users free of charge.

Our first Deliverable 1 sub-task (1.1) is for ABOS to create advertisements for the BOIR in *American Libraries* and *Public Library Association* magazines. The BOIR is a new initiative, and while they have the infrastructure to support data collection, they need national support and awareness for their work to succeed. The BOIR contains a limited amount of data, and nationally advertising the BOIR will speed up data collection. The BOIR tool has already been created and requires further adoption by public libraries. Sub-task 1.2 calls for ABOS to make ongoing usability enhancements to the BOIR. Starting in October 2022, ABOS will begin advertising the BOIR (1.3) for one year. Data collection (1.4), which is

already underway, will start at the same time but end July 2023. Another subtask (1.6) provides two years of website hosting for the BOIR project.

A key sub-task (1.9) is for ABOS and PLA to provide schemas that describe their datasets to ICPSR before depositing data for curation. Once ICPSR receives both schemas, PLA will provide an initial wave of data (sub-task 1.7) to ICPSR for processing. This preliminary task begins in October 2022 and ends in July 2023. Upon completing Deliverable 2, ICPSR will start curating BOIR data (1.5) and the second wave of PO data (1.8). ICPSR will continuously curate data to allow recursive, iterative changes before releasing the final dataset in 2024. We will save data in multiple formats (CSV, R, SAS, SPSS, STATA) and provide a standard codebook to maximize reuse.

The final product we will create is a one-time extract of available data, which could be augmented annually in future years. **We do not commit to future data releases because this is a pilot project.** However, if we are successful, we may approach IMLS to sponsor further data releases.

ICPSR staff responsible for Deliverable 1 include **A.J. Million**, **Rujuta Umarji**, and a to-be-named Data Curation Supervisor, Curator, and Project Manager. Million and Umarji will provide administrative support. A Project Manager will support Million while the Data Curation Supervisor and Curator will process data. **Sara Goek** from PLA and ABOS President **Lori Berezovsky** will ensure data are deposited at ICPSR on time using the resources described in our **Budget**. ABOS Board Member **Maggie Peterson** will coordinate BOIR data collection and tool enhancement. At the end of our project, librarians at ICPSR will scan published literature to identify bibliographic references to our dataset.

Data Curation Services: ICPSR provides a default suite of services for all of the data it collects, curates, and disseminates. Data at ICPSR comply with the FAIR principles—findability, accessibility, interoperability, and reusability. Curation includes a review of data for disclosure risk. If issues are found, data curators address confidentiality concerns (e.g., masking variables) to protect research subjects while maximizing the analytic potential of data. Curators create documentation and perform data cleaning tasks, including checking for undocumented codes and standardizing missing data. ICPSR's curation process includes, but is not limited to, data coding, data clean-up, file system management, metadata creation, and providing data in multiple file formats. We will:

- Review data for sensitivity
- Convert software-specific documentation to PDF
- Generate multiple data formats for dissemination and preservation
- Create a descriptive metadata record
- Assign a Digital Object Identifier (DOI), which recognizes the contribution of the researcher and ensures that the data are discoverable
- Check for undocumented/out of range codes
- Create variable labels
- Create value labels
- Identify and address foreign language characters
- Optimize file sizes
- Gather citations to related publications for the ICPSR Bibliography of Data-related Literature

Products of the curation process at ICPSR include a full suite of data files and comprehensive documentation offering study-level information and detailed variable information with displays of frequency distributions. All documentation is provided in PDF format and encoded in the Data Documentation Initiative (DDI) standard, an international metadata standard for describing social science data. ICPSR's processes align with the Open Archival Information System (OAIS) Reference Model, produced by the NASA Consultative Committee for Space Systems. OAIS provides the functional framework for sustaining digital objects in online repositories, supported by the International Organization for

Standardization (ISO). Curation deposits data in the ICPSR General Collection or a topical sub-archive. All ICPSR data are available online via a public website supplemented with a suite of search tools. ICPSR is also a CoreTrustSeal archive, meaning it archives data with an eye toward long-term preservation and use.

Deliverable 2: Our second deliverable is to create a data module. Our data module will inform data curation at ICPSR; link PLA and ABOS data to the PLS; and pilot a model for collecting, storing, and disseminating data about public library outreach and programming outcomes. Our module will be a technical report that maps variables from our dataset to the PLS by establishing a pre-defined unit of analysis. **Specifically, we will map individual libraries to their administrative units found in PLS data and specify upcoding transformations by ICPSR curators**. Mapping libraries to their administrative units will allow researchers to combine our data with the PLS because both datasets share a key variable. Our data module will also document variables we compute and those from third-party resources (e.g., demographics from the Census Bureau). For example, we may determine that broadband availability predicts programming and outreach outcomes, so we would add it to our dataset, explain why, and cite the Federal Communications Commission's (2020) Fixed Broadband Deployment Data as a source.

Our first Deliverable 2 sub-task (2.1) is to form an **advisory committee**. We will convene a committee in October 2022, and this group will be responsible for creating our data module. The committee will include 15 individuals representing project partners (ICPSR, SISLT, PLA, ABOS); urban, suburban, and rural public libraries; state libraries; a statistician from the University of Michigan; and historically underrepresented groups in the library profession. We will recruit committee members through an open, web-based call for participation and team member networks such as the National Associations of Librarians of Color. We will offer $250 gift cards to incentivize committee membership. Independently, the committee will establish an internal governance structure for building consensus and documenting action plans. **Denice Adkins**, Professor at the University of Missouri, an expert in public libraries and minority population outreach, will lead our committee.

After convening, the project committee will identify data transformations and enhancements (2.2) to improve our dataset. Committee members will lend their subject-matter expertise. The committee will map BOIR and PO datasets to the PLS (2.3) at this time, and the result will be a draft task list for ICPSR's Curation unit. Between March and July of 2023, the committee will write the data module (2.4), determine which layout it will take, and document how they created the module. Aside from providing technical information specifying details associated with our final data product, the module will provide a narrative documenting committee governance, challenges, and successes. The report will function as our pilot model for library programming, outreach, and outcome data governance, and we will publish it in July 2023.

Deliverable 3: Our final deliverable will be educational materials and outreach to promote the use of our data and share lessons learned with library stakeholders. We will take two approaches toward outreach and education:

1. SISLT-led outreach to students and educators
2. PLA-led outreach to state libraries, library administrators, and researchers

During the summer of June 2022, Adkins will work with a Graduate Student at the University of Missouri to develop open-access **learning objects** (e.g., lesson plans) that can be imported into library school graduate courses (3.1). At SISLT, learning objects will be tested in three classes: *Public Libraries*, *Library and Information Agency Management,* and *Research Methods*. Our learning objects will harness problem-based learning and incorporate design thinking where students are motivated to create solutions to ill-structured problems. For *Public Libraries*, a learning module will emphasize data-driven decision-making related to programming. A *Library and Information Management* learning object will emphasize coordinating statewide data collection related to declared priorities. A *Research Methods* learning object will focus on determining the impacts of types of programs on societal well-being in rural and urban areas.

We will share our learning objects via the ICPSR website alongside our data and module. After we create our learning objects, Adkins and Million will market them to ALA-accredited library programs (sub-task 3.2). They will also disseminate learning objects by presenting at the Association for Library and Information Science Education (ALISE) annual conference, Library Research Roundtable webinars, and advertising to LIS programs with support from the SISLT Graduate Student.

Parallel to SISLT creating learning objects, PLA will generate marketing and training materials for library administrators, state librarians, and researchers. **Sara Goek** will take the lead, and working with the staff at PLA, she will: 1) describe how to access our dataset and data module, 2) write a brief summarizing critical findings in our data, and 3) offer guidance for state libraries others who wish to replicate our governance approach (sub-task 3.3). PLA will also organize webinars and presentations free of charge. Once PLA creates training and outreach materials, it will begin sub-task 3.4, which calls for advertising digital products via social media, sending emails to PLA members, reaching out to state libraries, and other activities. ICPSR will support PLA through in-kind contributions by its Marketing and Communications Unit. See our **Digital Products Plan** for a list of the materials we will create and advertise.

Resources, Project Management, and Performance Metrics: We request **$45,053** for salaries and wages, **$13,518** for benefits; **$82,162** for indirect costs like facilities; **$9,947** for systems infrastructure to perpetually archive data; and **$3,750** for advisory committee incentives. We also request **$95,252** for ICPSR sub-awards and contracts with ABOS, PLA, and SISLT. Our **Budget** and **Budget Justification** provide detailed information about the resources we require. Million will oversee project management with ICPSR staff. However, to promote efficiency, tasks other than data curation will fall to ABOS, PLA, and SISLT. Adkins will convene the advisory committee and identify and suggest data enhancements with committee assistance. She will develop learning objects with the SISLT Graduate Student. Goek and a Program Coordinator will provide PO data to ICPSR, create outreach/training materials, and organize webinars. Berezovsky and Peterson will enhance the BOIR tool with help from a contractor and deliver data to ICPSR for curation. See our **Performance Management** Plan for how we expect to ensure our project will stay on target.

Community Feedback: This project is two years in the making, and ABOS and PLA representatives were involved in preliminary planning efforts. To guide project activity, we will hold monthly advisory committee meetings during project year one that includes ABOS, PLA, SISLT, and other community representatives. Adkins will set agendas to engage advisory committee members, and Million will update the committee on project activity at ICPSR. During our meetings, we will build consensus by asking committee members to rate data module features by using assessment scales. Diversity of opinion is a welcome benefit to allow ICPSR to see multiple end-user perspectives.

Diversity Plan: ICPSR recognizes diversity is closely connected to its mission, so the organization works to create an environment characterized by diversity, equity, and inclusion. This project explicitly meets ICPSR's goal of diversifying local data holdings. If our advisory committee recommends adding pertinent variables to our data, this project may shed light on diversity-related issues.

Our committee will also promote a diversity of perspectives when creating our data module. When we create our module, committee members will share their understandings of representative, localized community data needs. For example, a librarian representing the Latinx community may be interested in adding data about English as a second language to our dataset. Incorporating diverse perspectives will help us work toward **data equity** by collecting, archiving, and disseminating data without bias or exclusion. Our committee will also promote technical project diversity. We will seek individuals with diverse skills, job titles, and domain knowledge when recruiting committee members. Furthermore, we will structure the dataset we create to optimize equitable reuse by bringing together varied expertise.

Commitment to Diversity, Equity, and Inclusion: We will strengthen librarianship's commitment to diversity, equity, and inclusion by 1) maintaining a diverse advisory committee, 2) searching for ways to link BOIR and PO data to diversity-

related variables, 3) creating educational materials that emphasize how our data can promote equitable library services, 4) and piloting a collaborative, inclusive model of data governance. Because public library services vary based on localized community needs, we will attempt to factor this into our data by ingratiating contextual demographics. When advertising digital products, we will make significant efforts to work with diverse communities, potentially making custom presentations at conferences like the Joint Conference of Librarians of Color, REFORMA National Conference, the National Conference of African American Librarians, and the Association for Rural and Small Libraries.

Project Results: This project will aggregate, curate, and archive a dataset that documents the impact of library programming and outreach on communities, simultaneously prototyping a collaborative model to help address the lack of a national strategy for outcome data governance. We will advance knowledge and understanding in librarianship, maximizing federal dollars, by disseminating knowledge and educational resources created for this project. The knowledge we create will relate to insights gleaned from our newly created dataset and our experience developing a data module facilitating the creation of our dataset. Applied knowledge learned from how we aggregate and disseminate our data will be of value because it will provide a way to harness distributed data sources to supplement the PLS.

Potential benefits to society include librarians using our dataset to make practical service decisions. Our project will help library administrators make informed service decisions with limited resources; provide evidence for researchers to document the value of public library services; and support library outreach to stakeholders like municipal leaders, policymakers, and potential partner organizations. **Another benefit to society is the potential to produce better data at a lower cost by leveraging federal, state, and association contributions.** Suppose public libraries and professional associations work with IMLS and the Federal government to augment the PLS with an alternative data collection approach. In that case, there is significant potential for cost-savings, better data, and more information for national and local decision-making regarding the prioritization of public spending. This approach aligns with a broader shift in governance over the past 50 years from a traditional bureaucratic model to a collaborative process that optimizes the strengths and resources of diverse stakeholders.

Our data module will provide structure and support for future data projects to ensure project deliverables are adaptable, generalizable, and usable elsewhere. We will also curate data in standardized formats, readily adaptable to local environments. Regarding our data module, we will provide it in an easy-to-read PDF document that maps variables from BOIR and PO data to the PLS, documents data transformations that we conduct, and identifies the sources of additional data points and variables we included in the data. This data module will guide researchers to draw inferences from our data in the context of the PLS, and it will also contain information describing how we created it. To ensure lessons from our study are adaptable and usable by others, we will document lessons learned and provide a summary of activities in our data module alongside best practices we find.

Because members of the library community may lack technical skills to analyze our data, PLA will also develop1 a data brief that contains descriptive statistics. See our **Data Management Plan** and **Digital Products Plan** for more information about our plans to document and preserve project knowledge.

How We Sustain Project Benefits: We will sustain our project contributions to society by 1) **archiving data** and 2) **providing clear guidance to public libraries about how to collect and disseminate data demonstrating the impact they have in their communities**. ICPSR will perpetually underwrite public storage of our dataset and module to sustain benefits derived from others using our dataset. ICPSR also participates in community initiatives that preserve data for coming generations, including the Digital Preservation Network (DPN). DPN provides members of the academy and their successors with the assurance that future access to their resources will be available in a disruptive change in administrative or physical institutional environments. ICPSR also maintains access to special-purpose, cloud-based infrastructure to support its digital preservation mission.

Aside from ICPSR storing data for reuse by librarians, we will provide access via a catalog with a permanent DOI and offer user support to researchers, students, policymakers, policy analysts, practitioners, the media, and the public. ICPSR maintains a YouTube channel with videos, tutorials, step-by-step instructions, and demonstrations that guide data use. Data users at ICPSR member schools may also contact their local ICPSR representatives for assistance in finding, accessing, and working with ICPSR data. ICPSR's user support team commits to providing user support, vetting restricted data access requests, and continuously helping with end-user questions.

More generally, we will sustain project benefits by creating a pilot data model for collecting, storing, and then disseminating public library data and service outcomes. If our project is successful, we anticipate pursuing additional funding to curate new data; however, in the event we do not, another way we will sustain the benefit(s) of this study is by providing training, a data module, and other resources to enable the library community to undertake similar initiatives. In short, we will build out the technical and organizational infrastructure necessary for collective action within the library community to measure impact while using dispersed resources from government, higher education, professional associations, and public libraries. By doing so, we will ensure benefits will be felt from our project for years to come. If our data curation and dissemination model is ineffective, we will rule out a national direction for public libraries to collect and analyze data to demonstrate their impact.

Task	Activity	Year 1											
		2022-2023											
		Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
Deliverable 1													
1.1	Design advertisements	X	X										
1.2	Enhance BOIR tool capacity to ingest data			X	X	X	X						
1.3	Advertise tool			X	X	X	X	X	X	X	X	X	X
1.4	Expedited BOIR data collection			X	X	X	X	X	X	X	X	X	X
1.5	BOIR data delivered to ICPSR and curated			X	X	X	X	X	X	X	X	X	X
1.6	Host BOIR tool	X	X	X	X	X	X	X	X	X	X	X	X
1.7	Wave 1 - PO data deposited at ICPSR and curated			X	X	X	X	X	X	X	X	X	X
1.8	Wave 2 - PO data deposited at ICPSR and curated												
1.9	Data schemas provided to ICPSR	X											
Deliverable 2													
2.1	Form advisory committee	X	X										
2.2	Establish pre-defined units of analysis, variables of interest, and conduct data transformations			X	X	X	X	X					
2.3	Map BOIR and PO schemas with PLS			X	X	X	X	X					
2.4	Create data module								X	X	X	X	X
Deliverable 3													
3.1	SISLT-led materials created											X	X
3.2	SISLT-led materials disseminated to ALA-accredited LIS programs												
3.3	PLA-led materials created												
3.4	PLA-led materials to state libraries and practitioners												

Task	Activity	Year 2											
		2023-2024											
		Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
Deliverable 1													
1.1	Design advertisements												
1.2	Enhance BOIR tool capacity to ingest data												
1.3	Advertise tool	▓	▓										
1.4	Expedited BOIR data collection												
1.5	BOIR data delivered to ICPSR and curated	▓	▓	▓	▓	▓	▓	▓	▓	▓			
1.6	Host BOIR tool	▓	▓	▓	▓	▓	▓	▓			▓	▓	▓
1.7	Wave 1 - PO data deposited at ICPSR and curated			▓	▓	▓	▓	▓					
1.8	Wave 2 - PO data deposited at ICPSR and curated			▓	▓	▓	▓	▓	▓	▓			
1.9	Data schemas provided to ICPSR												
Deliverable 2													
2.1	Form advisory committee												
2.2	Establish pre-defined units of analysis, variables of interest, and conduct data transformations												
2.3	Map BOIR and PO schemas with PLS												
2.4	Create data module												
Deliverable 3													
3.1	SISLT-led materials created	▓											
3.2	SISLT-led materials disseminated to ALA-accredited LIS programs		▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓
3.3	PLA-led materials created						▓	▓					
3.4	PLA-led materials to state libraries and practitioners								▓	▓	▓	▓	▓

Overview: We anticipate creating a variety of digital products for this project. We will manage and provide these products to members of the library community, academic researchers, graduate students, and the general public. Below we describe the types of digital products we will create, their availability, how products will be accessible, and the sustainability of our products.

Type: We will create three types of digital products: a **dataset**, a **data module**, and **educational and outreach resources**. Our dataset will include data from the Public Library Association's Project Outcome (PO) and the Association of Bookmobile and Outreach Services' Bookmobile and Outreach Information Repository (BOIR). Our dataset will also contain variables from sources like the U.S. Census Bureau or entirely new variables that we compute. Our data module will be a report that draws linkages between our dataset and the Institute for Museum and Library Services' Public Library Survey (PLS) by connecting variables at the library administrative unit level. Our module will also describe how we created our dataset and help other organizations partner with multiple library stakeholders to aggregate and disseminate national data. Finally, the educational and outreach resources we build will include open-access modules for library graduate courses, a brief summarizing our data, materials describing how to access our dataset, and other outreach content.

We will provide all data files in standard formats like ASCII, SAS, SPSS, Stata, and R. Our data documentation like codebooks will be provided in PDF format. Our data module will be in PDF format. All outreach and communication files will be saved as PDF, PPTX, JPEG, GIF, or MP4.

Availability: All digital products we create will be freely available to the public. Our final, curated dataset will be archived at the Inter-university Consortium for Political and Social Research (ICPSR) and provided to users via restricted download. Restricted download requires researchers to apply to use data through our online catalog, provide us with clear reasons for using our data, and pledge to maintain the confidentiality of identifiable parties. We will host our data module on the ICPSR website with other usage restrictions. Finally, we will provide our educational and outreach resources on the PLA and ICPSR websites. All our content will be provided under a Creative Commons Share-alike license (CC BY-SA).

Access: After providing data to ICPSR, ABOS and PLA will sign a deposit agreement that defines their rights as data providers. ICPSR will be responsible for providing perpetual access to data and protecting the privacy and confidentiality of individuals and organizations. We will provide data access via restricted download because we anticipate that information regarding individual libraries will be in the data. We do not expect any risk to libraries or patrons associated with others sharing our data. However, we will not allow users to redistribute our data to respect libraries' data sovereignty. Last, we will provide free online access to our data module and our educational and outreach resources.

Sustainability: ICPSR is a CoreTrustSeal, trustworthy data repository. CoreTrustSeal offers to any interested repository a certification under the umbrella of the Research Data Alliance to merge their data repositories' certifications. Certification requirements facilitate long-term sustainability and establish succession plans if archives close or cannot function. ICPSR has over six decades of experience preserving data and adapting to technological change. An essential practice is ensuring redundancy of stored data, which ICPSR achieves by using multiple and varied methods and locations to back up its holdings.

ICPSR also participates in several initiatives that focus on preserving data for future generations, including the Digital Preservation Network (DPN), the only large-scale digital preservation service built to last beyond individual, technological system, and organizational life spans. DPN provides members of the academy with an assurance that future access to their scholarly resources will be available in the event of disruptive change in administrative or physical institutional environments. ICPSR also uses the DuraCloud service and Fedora software, which is both community-maintained and open-source.

Data Description: Data for our project comes from the Public Library Association (PLA) and the Association of Bookmobile Outreach Services (ABOS). PLA's Project Outcome (PO) initiative provides an online tool with standard survey instruments for public libraries to measure patron attitudes about library programs. More than **2,000** libraries have created surveys, and patrons have responded **356,940** times. Library survey data provide a snapshot of public library civic engagement, early childhood literacy, economic development, education, and public health programming outcomes. ABOS' Bookmobile and Outreach Information Repository (BOIR) collects data from public libraries about library bookmobiles and outreach programs. BOIR data is valuable because it provides administrative data about public libraries' outreach services (i.e., *inputs* and *outputs*).

Responsibility: PLA and ABOS will provide data to the Inter-university Consortium for Political and Social Research (ICPSR). ICPSR will ingest, curate, describe, disseminate, and preserve BOIR and PO data.

Designated Archive: The ICPSR General Archive is the world's largest digital social science data archive. It contains 16,000+ studies, 250,000 files of research, and 21 specialized topical data collections. **This integrated data management plan leverages the capabilities of ICPSR and its trained archival staff.**

Access and Sharing: ICPSR will share data with academic researchers, state librarians, the general public, and library administrators.

- *Public-use Data Files*. These files have no direct and indirect identifiers, and we will provide access online through the ICPSR website. After users agree to our **Terms of Use**, anyone with a MyData account may download them.
- *Restricted-use Data Files*. These files contain identifiable and disclosive information. Users must apply to use these files, create data security plans, and agree to other access controls. We anticipate most project data will be restricted use, and we will retain identifiable information to facilitate analysis by users.
- *Timeliness*. ABOS and PLA will promptly supply ICPSR with the data.

Selection and Retention: ICPSR will archive ABOS and PLA data and associated documentation. ICPSR will curate the data to adapt to changing technologies over time.

Metadata: ICPSR will create substantive metadata in compliance with the most relevant standard for the social, behavioral, and economic sciences—the Data Documentation Initiative (DDI). This XML standard provides for the tagging of content, which facilitates preservation. We will create five types of metadata:

- *Study-Level Metadata Record*. We will create a study-level DDI record for the ICPSR catalog then index it with key terms from the ICPSR Thesaurus to enhance discovery.
- *Data Citation with Digital Object Identifier (DOI)*. IPCSR will create a dataset DOI to permanently identify data and ensure they are always accessible.
- *Variable-level Documentation*. ICPSR will tag variable-level information in ICPSR's variable database.
- *Technical Documentation*. The variable-level files described above will serve as the foundation for documentation in a codebook that ICPSR will prepare and deliver.
- *Related Publications*. ICPSR will search for publications that cite ABOS and PLA data and provide two-way linkages between data and publications.

Intellectual Property Rights: Creators and their institutions own the data they generate. By depositing with ICPSR, investigators do not transfer copyright but grant permission for ICPSR to disseminate the data and transform it to protect respondent confidentiality, improve usefulness, and facilitate preservation.

Ethics and Privacy: After ABOS and PLA deposit data, it will undergo procedures to protect the confidentiality of organizations and individuals. These include:

1. Review to assess disclosure risk
2. Modifying data as necessary to protect individual and organizational confidentiality
3. Limiting access to datasets in which the risk of disclosure remains high
4. Consulting with ABOS and BOIR to manage disclosure risk

ICPSR will assign a data curator certified in disclosure risk management to act as a steward for data while they are processed. Data will be processed and managed in a secure, virtual environment.

Format: ABOS and PLA will submit data to ICPSR in plain text format. Furthermore, they will provide data documentation in PDF or Word. ICPSR will make data files accessible in multiple standard formats (e.g., ASCII, SPSS, Stata, R) and documentation (e.g., codebooks) in PDF. ICPSR will store master files using prevailing standards. ICPSR currently keeps data in ASCII format and stores documentation using XML and PDF/A.

Archiving and Preservation: ICPSR is a data archive with a 60-year record of preserving and making research data available over several generational technological shifts. ICPSR accepts responsibility for the ongoing preservation of all research data upon receipt of a signed deposit form. This responsibility includes an open commitment to managing successive iterations of the data if new waves or versions are deposited. ICPSR will ensure the research data are migrated to new formats, platforms, and storage as required by good practice in the digital preservation community. Digital preservation requires an organization to address succession planning, and ICPSR will designate a successor in the unlikely event that such a need arises.

Storage and Backup: Research has shown that multiple locally and geographically distributed copies of digital files are required to keep information safe. Accordingly, ICPSR will place a master copy of each digital file (i.e., data files, documentation, and other related files) in ICPSR's Archival Storage, with several copies stored with partner organizations at designated locations and synchronized with the master.

Mission Statement: The Inter-university Consortium for Political and Social Research (ICPSR) advances and expands social and behavioral research, acting as a leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations.[1, 2]

About: An international consortium of more than 750 members, ICPSR is the largest digital archive of social and behavioral science data in the world. ICPSR curates, archives, and disseminates social science data while providing leadership and training in data access, curation, and analysis methods for the social science research community. In 2019, ICPSR was awarded a National Medal for Museum and Library Service. More recently, ICPSR received a $38 million commitment from the National Science Foundation to create a 21st-century data environment to help researchers access, collect, store and secure vital information.

ICPSR maintains a **data archive** containing more than 16,000 data collections with over 250,000 data and documentation files. ICPSR collaborates with funders, including U.S. statistical agencies and foundations, to create thematic data collections, encourage data stewardship, and conduct research projects. ICPSR provides social scientists and others with **educational activities** like the Summer Program in Quantitative Methods of Social Research, a comprehensive curriculum of intensive courses in research design, statistics, data analysis, and social methodology. ICPSR **research** focuses on cutting-edge challenges related to digital curation and data science. ICPSR leads and participates in policy initiatives and grant-funded activities that result in publications that address data stewardship issues. ICPSR researchers also examine other issues related to topical data collections.

Organization and Governance: ICPSR is a center within the Institute for Social Research (ISR) at the University of Michigan. ISR is the world's largest academic, social science survey and research organization and a leader in creating and applying new social scientific methods to research. A public research university, the University of Michigan, was established in 1817. In the 2021 edition of the U.S. News and World Report "National University Rankings," the university was 23rd among national universities.

ICPSR is governed by a 12-person Council elected by the ICPSR membership. Council members are subject matter experts in their fields and represent key substantive social and behavioral science research areas. Members serve four-year terms, and six new members are elected every two years. The Council acts on administrative, budgetary, and organizational issues on behalf of ICPSR members. ICPSR governance takes place within five standing committees: Budget and Policy, Collection Development, Membership and Marketing, Training and Instruction, and Preservation and Access. ICPSR follows three documents: a Constitution, its Bylaws, and a Memorandum of Agreement with the University of Michigan and ISR.

ICPSR's funding model offers stability and reliability. Member institutions contribute to ICPSR, including nearly every major research university in the U.S., teaching colleges, research organizations, and universities in other countries. The faculty, staff, and students at member institutions receive access to the data archive and related services. Supplementing the services ICPSR provides to its members, it also serves 21 government agencies and foundations, including the Department of Justice, Bureau of Justice Statistics; the Bill and Melinda Gates Foundation; the Substance Abuse and Mental Health Services Administration; and the Robert Wood Johnson Foundation.

---

[1] ICPSR. (2021, May 1). *About ICPSR*. About the Organization. Retrieved February 7, 2022, from https://www.icpsr.umich.edu/web/pages/about/

[2] Approved by the ICPSR Governing Council.