

Narrative

Introduction

The Data Storytelling Toolkit for Libraries (DSTL) connects two visions: data storytelling training from the University of Illinois at Urbana Champaign (UIUC) School of Information Sciences (iSchool); and library data tools and infrastructure from the Public Library Association (PLA). The DSTL will provide data storytelling guidance and tools that help libraries reach audiences with stories based on data. This project meets program **Goal 3** enhance the training and professional development of the library and archival workforce; **Objective 3.1** develop training to equip the library and archival workforce to engage in sustained community development through data analytics and communicating collective impact; and **Objective 3.3** create and/or refine training programs to build library and archival workforce skills, with a particular focus on cultivating creative skills and self-directed learning opportunities for workers.

Storytelling is woven deep into human neurology; narrative experience interconnects individuals and stories are a fundamental source of human information.¹ In the LIS tradition, storytelling is a dynamic exchange between the teller, audience, and the story.² Captivating an audience of 90+ bouncing preschoolers would be daunting to most CEOs, and yet librarians navigate this gauntlet every day.³ With proper training in storytelling, all library staff can learn how to captivate audiences using narrative structures and strategies. This is a new opportunity in the field, combining the over-130-year history of storytelling practice in public libraries with data storytelling. **Data storytelling** means any communication of data in story form, to make accurate information memorable and meaningful.⁴ Libraries have access to more data tools than ever. The challenge they face is marshaling the available evidence into meaningful stories that resonate with key stakeholders and demonstrate the value and impact of library services for their communities. There is a gap between available data and usable stories. The DSTL is the first project to address that gap by guiding users to develop skills to translate data into evidence-based stories.



Imagine a library that serves a town of about 30,000 people in the southwestern United States. Staff have noticed that literacy programs for children have remained popular, but more people are asking for Spanish language materials. At the same time, they notice a decrease in program attendance. Library staff begin to examine library data—numbers of programs, attendance at programs by ages, children’s circulation—and community demographic data—census data, school district data. They see that their community has undergone a demographic shift in the past decade, with an increase in the Spanish-speaking population. The library needs to better understand and serve these community members.

Library staff deserve easily accessible online support for generating stories from data. The DSTL will connect tools for structuring stories directly to real-world inspirational examples and data, addressing the gap left between data projects and published library impact stories. Library staff will be able to learn and engage through two toolkit pathways: create and explore.

- The **create** pathway will walk users through the development of a story from data they upload. This builds on the demo created for the DSTL planning grant (RE-250094-OLS-21).

* See Supporting document 1: References for all citations.

Implementing the Data Storytelling Toolkit for Librarians

- The **explore** pathway will guide users to data story templates based on pre-existing data from their public library in Benchmark: Library Metrics and Trends (librarybenchmark.org). Benchmark is a joint initiative of PLA and the Association of College and Research Libraries (ACRL).⁵ For public libraries it provides easy access to data from the IMLS Public Libraries Survey (PLS),⁶ PLA annual surveys, and the Census Bureau’s American Community Survey in interactive dashboards with visualizations focused on peer comparisons. By building a pathway of the DSTL into Benchmark we take advantage of its developed infrastructure and ensure the toolkit can be updated and sustained in the future.

Both pathways will be interactive and will integrate resources for hands-on skill-building. DSTL users will be able to find specific resources to meet a need—for example, how to reach challenging audiences—or work in sequence through the step-by-step process to generate a data story targeted at those audiences. The development of the toolkit will occur in four phases:

Project Phase	Outcomes
1. Project Initiation	Planning and hiring Library recruitment for interactive design
2. Research and Design	Create pathway design and iteration Explore pathway design and iteration
3. Production and Testing	Implementation of both pathways Testing and evaluation
4. Finalize and Promote	Ensuring access and sustainability Dissemination, publication, and conferences

The storytelling guidance will be informed by the PIs robust research. It will help users learn how to select from classic goals and motivations and understand strategies for reaching audiences and structuring narratives. With the DSTL, library staff can train themselves, combining effective story structures and data visualizations to tell their own compelling library impact stories.

Project Justification

Needs and Challenges

The DSTL seeks to overcome barriers to library data storytelling and inspire both story experts and data experts. A recent study of Wisconsin public library staff revealed that while the majority use data at least monthly, there are still significant barriers. Most survey respondents were untrained in data use: “Over 70% of survey respondents had no professional development or continuing education related to data and, of those people, 53% also had no formal academic training related to data work (37% of all survey takers).”⁷ Gaps in knowledge can lead to a lack of confidence. The study found that data are mostly collected and analyzed to meet basic reporting requirements, and public library staff would benefit from additional resources like dashboards and templates, as well as training opportunities where participants could use their own libraries’ data to build skills and confidence. Recent qualitative research confirms these findings, revealing two kinds of obstacles that inhibit public library data storytelling: both fears of data complexity and story-crafting accuracy; and lack of time, tools, and training resources.⁸

Existing initiatives have developed a robust data ecosystem for public libraries. These efforts include PLA’s Project Outcome, a comprehensive resource for outcome-based evaluation that is very useful for assessment of programs and services but does not address all areas of library impact. They also

Implementing the Data Storytelling Toolkit for Librarians

include the Research Institute for Public Libraries (RIPL), a hands-on experiential learning immersion for measuring library impact, which is limited by its periodic availability to slightly over 100 applicants and an in-person-only format.⁹ The Measures that Matter (MtM) group,⁶ with participants who work with data as experts, leaves out the many novices who are uninformed about or even skeptical or fearful of data. Finally, launched in fall 2021, Benchmark: Library Metrics and Trends is a joint initiative of PLA and the Association of College and Research Libraries (ACRL).¹⁰ A 2023 survey of Benchmark users found that they primarily use the tool for strategic planning, decision making, and reporting to stakeholders. While the respondents (72% administrators and 21% data analysts or evaluators) indicated that they find the data useful for peer comparisons and prioritizing efforts, they want to be able to better engage all library staff in that process and they feel they need additional training to make effective use of data.¹¹ Integrating the DSTL with Benchmark will address those needs.

Library impact stories are available from subscription journals like *Library Journal*, press releases from the American Library Association, in library or system newsletters and other reports. But these stories are not always positioned to inspire further storytelling. For example, the Library Services and Technology Act (LSTA)¹² reports are rich sources of success stories with new keyword searching. But successful search requires knowing that this is a stories repository, which is not clear, as well as keywords for describing library impact, which may be challenging for new professionals. There is no controlled vocabulary for library success stories, which lack access points beyond online search.

The first step in meeting these challenges was the 2022-2024 DSTL planning grant that developed the basic model for this project with input from over one hundred library workers, including directors, data analysts, communications specialists, and frontline staff (see supporting document 2: interim performance report). The planning grant launched in fall 2022 with a workshop that saw just over 680 registrants, and [the publicly available webinar recording](#)¹³ has had almost 200 views since then. Thanks to this IMLS support, we have produced a basic design for one pathway (create) with an [interactive demo](#) for crafting data stories. Every aspect has been tested by real library staff and iteratively redesigned based on their input.¹⁴ Beyond this project, 400 people attended the PLA webinar on "[Data Storytelling: Advocating for the Library and Community](#)."¹⁵ Data storytelling for libraries is in demand.

Target Audience and Beneficiaries

The DSTL will be designed so that even the busiest library directors who single-handedly run small libraries can easily use and benefit from data story guidance. This commitment is important because the majority of US public libraries are in town or rural areas (70% of administrative entities) and many have limited staff capacity.¹⁶ Specialized data analysts in large public libraries and systems will also benefit, gaining strategies for building narratives and reaching audiences with their analyses. At the UIUC, LIS application essays from the past 20 years show that many are drawn to the profession by their love of reading, and 93% come from humanities disciplines.¹⁷ Through the DSTL such story experts will gain motivation to engage with data visualization, exploring their public library's data in Benchmark or creating a narrative based on their own data. Library staff will build data literacy skills while learning to communicate memorably and meaningfully in story forms so that they can serve changing populations, for example, supporting literacy for children and families who speak several languages.

Building on Existing Scholarship and Practice

The DSTL differs from other projects by combining 1) the stability of the established library data portal Benchmark: Library Metrics and Trends with 2) innovative data storytelling instruction and 3) lessons learned from a successful planning grant.

First, this project requires partnership between a university and a professional organization with an already recognized initiative—PLA Benchmark. PLA has a successful track record of managing projects of a similar scope, so creating and publicizing the DSTL will be more efficient and effective than continuing to develop a standalone tool from the planning grant. PLA will leverage ongoing relationships with developers who have built nationally significant tools, including Project Outcome and Benchmark. Currently, Benchmark allows users to explore their data through a suite of interactive data dashboards and visualizations. However, PLA has found that users still struggle to apply what they find in practice.¹⁸ The DSTL will help library staff move from simply collecting or finding data to using that data to communicate more effectively with their stakeholders.

Second, this project leverages successful data storytelling courses, which appear to be the first of their kind in any field, taught at the iSchool at Illinois since 2017. These courses bring together expertise in educational approaches for professionals in the library workforce for over 125 years. Data storytelling instructors have 58 combined years of experience in teaching storytelling and data visualization. This work has contributed to advances in the theory and practice of teaching data literacy.¹⁹ Courses are sustained by a team of four full-time faculty, two adjunct faculty, and a changing team of doctoral students with expertise in storytelling, library data, and data visualization. Dr. McDowell has conducted nine years of research in story as a fundamental information form²⁰ and storytelling as information practice—90 interviews, focus groups, workshops, questionnaires, and more. With Dr. Turk and RA Xinhui Hu, she co-developed personas for storytelling toolkit design²¹ that will inform the DSTL implementation process. We have the leading experts in library data storytelling training on this team.

1 Choose your data storytelling Motivation

Achievement Benchmarking Justification

2 Identify your data storytelling Goals

- Understanding Real Needs
- Serving Real Needs
- Building on Strengths
- Address Deficits

3 Benchmark For Serving Real Needs

Use evidence of needs met through services to compare with outside standards or trends.

4 Data Visualization Recommendations

General Guidance

Justification: Consider using bar or line charts to show how the previous need and how that need has been served. An arrow can draw the audience's attention to the magnitude of the change.

Data Visualization Samples

Event Impact

Show the impact with metrics by-month data

Month	Impact
Jan	28
Feb	26
Mar	12
Apr	18
May	8
Jun	5

- Screenshots from the DSTL demo
1. Choose a motivation for storytelling
 2. Identify the library's data storytelling goals
 3. Sample guidance based on the motivation and goals selected
 4. Data visualization recommendations based on the motivation and goals selected

Third, a successful UIUC-based planning grant (RE-250094-OLS-21) began in 2022 and has resulted in a functional demo (pictured above). The DSTL approach is distinct from other training and education programs because it enables staff with any level of data expertise to educate themselves, interactively choosing the pathway and guidance most relevant to their communication needs. Complementary create and explore paths - described below - will reinforce story ideas with data and data insights with story, providing self-guided learning and adaptable data story templates for all library staff.

The planning grant has established a track record of talks and publications.²² The DSTL team has put in place additional collaborations with RIPL and WebJunction at OCLC to publicize this work. Dr. McDowell was a speaker at RIPL in summer 2023 and has been invited to present DSTL insights at upcoming training sessions. Plans are underway for a Data Storytelling 101 webinar with WebJunction for June 2024 which will be recorded and added to their catalog, and additional webinars are under discussion. In the implementation grant, we will continue to build relationships with these partners and publicize the DSTL through webinars, talks, and conferences. PLA will host a free webinar to promote the DSTL widely to the field. Dr. Kate McDowell will be the keynote speaker at ALA's Library Research Seminar VIII as one of three keynote invitations fall 2024—every talk she gives publicizes the project. The DSTL team is well-positioned to support widespread adoption across the field.



*The library staff use the DSTL to develop a data story about serving community literacy needs to make the case for investment in more Spanish-language programming and books. Through the **explore** pathway, they clarify their goals and motivations, finding that their library offers fewer programs in both number and type relative to their peers and the types of programs they do offer are not currently well-matched to their community's demographic profile. The current literacy programs for children offered at the library are mostly in English, with only 2% in Spanish in the last year. In children's circulation records, they find a 10% increase in circulation of Spanish language materials without a commensurate increase in acquisition budget. The DSTL explore pathway guides them to use this data to make the case for what the library could do with a funding increase.*

*Through the **create** pathway, they learn more about strategies suited to communicating with their audiences, including the board, friends of the library, potential community partner organizations, the city council, and the public. These audiences have mixed attitudes towards the library currently and staff have not always had success in asking for additional resources. Based on what they learn from the DSTL, library staff decide they will try combining information and emotion in a classic discovery narrative, to communicate their research process and what they found about community needs and current offerings.*

Project Work Plan

The project work plan is designed to develop a toolkit that will comprise two pathways for guiding library staff to data stories tailored to their needs. With the **create** pathway, library staff create a story template by identifying their audiences, motivation, and narrative strategies, with the option to build a data visualization from among 40 categories of library data, such as program attendance and website visits. Like a choose-your-own-adventure book, each choice will generate a different set of data storytelling guidance. With the **explore** pathway, library staff choose from four key topics for which

Implementing the Data Storytelling Toolkit for Librarians

nationally significant data about public libraries exists (digital equity, library facilities, staffing, literacy). They choose the type of motivation and narrative they want to use based on guidance and end up with a dynamic data story that incorporates text and visualizations and shows how their library compares to peers. Both pathways will provide visualizations, story models, and guidance for library staff. Each pathway allows libraries to leverage insights from different types of data (see the Digital Products Plan for further details).

DSTL Pathways				
	Data	Visualization	Story	Endpoint
Create	Local library data (not stored or collected)	Upload option for data with downloadable visualization suggestions	Audiences, motivation, narrative structures, and optional data visualization	Guide for crafting a compelling story accompanied by data
Explore	Pre-loaded library data with peer comparisons	Local, national, and peer data visualized for immediate use	Narrative structures, arguments, and data as evidence	Downloadable data story with visualizations

The DSTL **create** pathway will be developed based on best practices in open-source coding, data privacy, digital access, and qualitative research. It will be built on open-source code, available in a public repository. UIUC plans to have no server-side computation, just client-side, to minimize website hosting demands. Design will employ an iterative process of public invitation for feedback. **Create** will be freely available to all libraries everywhere throughout the development process. It will include four sections that users may explore in any order: 1) reach audiences, 2) choose motivations and goals, 3) structure narratives, and 4) visualize data. These sections provide guidance and interaction for users to create any data stories that would benefit their library work.

The DSTL **explore** pathway will be implemented in Benchmark. It will go through a rigorous process of quality control characteristic of all PLA initiatives as part of the standard investment in excellent service to its members. Iterative design will be based on input from the Core Design Team and other Benchmark users. **Explore** will be integrated into a new interactive “Stories” section of *Benchmark*. It will guide users through a sequence of selections: 1) data/evidence related to a key topic, 2) motivation for storytelling, and 3) narrative approach. The user will end at a story model, pre-populated with their library’s data, that they can export and adapt to their needs. The topics will be organized around four public library core service areas that are of high importance to the field, with data available from multiple sources. PLA’s annual survey data²³ will be complemented by output measures collected via the annual IMLS Public Libraries Survey (PLS), and community data drawn from the Census Bureau’s American Community Survey. Provisionally, the four key topics we plan to use are digital equity, library facilities, staffing, and literacy. These will be confirmed through the iterative design process. See Attachment 3 for a concept map of the Explore pathway user experience.

The overall project team will include library data and storytelling experts from UIUC and PLA to coordinate planning, management, and execution. The UIUC team will lead the **create** pathway development, building on the planning grant demo with more rigorous testing with library audiences, iterative design, and implementation on a website hosted by the university. The PLA team will lead the

Implementing the Data Storytelling Toolkit for Librarians

explore pathway development, building on established data exploration and visualization available to libraries through Benchmark and incorporating guidance on narrative structures, arguments, and data as evidence. Both teams will benefit from regular coordination meetings to share insights, lessons learned, and align the design of the two pathways as one visually aligned toolkit. Project personnel include nationally recognized experts in data storytelling (McDowell), library data and evaluation (Goek and Jones), and data visualization (Naiman and Turk). McDowell, Naiman, and Turk at UIUC will develop the create pathway, and McDowell will supervise RAs. Goek and Jones at PLA will lead development of the explore pathway and evaluation of the toolkit. This team is one of the strongest nationally in training library staff to maximize the usefulness of data.

At UIUC, two students will be hired to lead iterative design (Design RA) and programming (Coding RA) respectively. The Design RA will contribute to all aspects of graphic design, interface design, and implementing qualitative research protocols designed by PIs for testing and evaluation toward iterative improvement of the DSTL. The Coding RA will focus on website architecture, including programming interactive data visualizations which will require learning open-source data visualization tools and techniques for the **create** pathway and using Power BI for the **explore** pathway. Both the Design and Coding RAs will be needed to effectively implement improvements identified through qualitative research and rigorous testing. These students will receive training in data storytelling and work with PLA as project partner for real-world experience. Additional students will benefit from the project through research practica or internships for course credit.

The planning grant relied on free website hosting and visual design to simulate online interactivity. In order to create a sustainable, robust DSTL that will meet national needs and become a reliable resource for all public libraries, the implementation project requires a larger investment in technical infrastructure, design, and programming expertise. This means significant time and financial support. The DSTL will be sustained beyond the period of performance through strategic investments in stability and technical infrastructure. For the **explore** pathway, Benchmark is a high-stability platform, with long-term investments in service to public libraries, and the robust development process will enable data storytelling guidance to become a permanent part of this tool and one that can be kept up to date. As outlined in the Digital Products Plan, Benchmark offers both free and subscription-based levels of access. The subscriptions are necessary to sustain its technical infrastructure. Further enhancing its usability requires investments beyond the current funding model. However, the benefits to public librarianship will be significant, and building out these functionalities within Benchmark will ensure the DSTL remains relevant and usable in the long term. For the **create** pathway, using open-source code will allow the UIUC-based website portion of the toolkit to be updated and revised over time. Ongoing revision is built into the process of course updating, since the DSTL will serve as a teaching tool for students in the information fields broadly, including LIS. Further details are included in the Digital Products Plan.

Schedule

The numbers used below correspond to tasks in the Schedule of Completion (Gantt chart).

In **Year 1 (2024-25)**, we will **initiate** the project and begin the **research and design** of two interactive pathways.

1.1: When the grant begins in August 2024, we will organize a collaborative system for teams across UIUC and PLA and set regular, recurring project team meetings.

Implementing the Data Storytelling Toolkit for Librarians

1.2: From fall 2024 through winter 2025 we will recruit 10 public libraries to the Core Design Team to provide regular, ongoing feedback on the iterative design process (PLA and UIUC) and conduct initial design workshops.

1.3: UIUC will confirm hiring of the Design and Coding RAs, who will begin work in the spring semester 2025.

2.1: PLA staff (Goek), in connection with the rest of the project team, will design and model two (first two of four) **explore** topics, digital equity and library facilities. We will work with Benchmark developers and the Coding RA to develop the technical functionalities and specifications.

2.2: In spring 2025, UIUC project staff (McDowell, Turk, and Naiman) will establish a University-based technical home for the **create** pathway based on prior existing and continued open-source projects that have focused on sustainability and peer-production. The Coding RA will begin programming work to implement the initial design of the create pathway, based on the DSTL planning grant. IRB approval for future focus-group evaluations will be obtained through UIUC.

The summer of 2025 into the beginning of **year 2 (2025-26)** will see the initial **production and testing** of the create and explore pathways.

3.1: The **create** pathway will be implemented in summer 2025. Once launched, UIUC will hold a public webinar to invite feedback, followed by invitations to focus groups and interviews (coordinated by the UIUC Design RA and Coding RA).

3.2: The first two topics in the **explore** pathway will launch in Benchmark by fall 2025. PLA staff (Goek and Jones) will gather feedback from the Core Design Team as well as other Benchmark users.

3.3: Concurrent with this process, PLA staff (Goek) will complete development of the other two topics in the explore pathway and work with developers to implement them within Benchmark early in 2026.

3.4: In spring 2026, project team (UIUC and PLA) will compile all feedback gathered via the various evaluation methods and incorporate it into feasible recommendations for improvements to the design.

2.4 / 3.5: Throughout year 2, the project team (UIUC/PLA) will engage in sustainability planning for the DSTL, considering factors including systems, expertise, operating infrastructure, and more as needed. These will be incorporated into the technical design by the end of year 2.

In **year 3 (2026-27)** we will implement the recommendations gathered through the iterative design process to **finalize and promote** the DSTL.

4.1: In fall 2026, the project team (PLA/UIUC) will gather additional feedback on both the create and explore pathways to evaluate their effectiveness and whether they will meet library user needs.

4.2: The improvements identified through user feedback will be implemented in both the create and explore pathways by spring 2027, resulting in the final version of the DSTL.

4.3: Final changes will include a web accessibility audit to ensure compliance with W3C guidelines and best practices for data visualization.

4.4: In spring 2027, the project team (PLA/UIUC) will work to implement the long-term sustainability plan through strategic planning, and assessing and improving alignment with PLA's service mission and UIUC teaching and research mission.

Implementing the Data Storytelling Toolkit for Librarians

4.5: The final evaluation (led by Jones at PLA) will consider the project's impacts overall and gather additional evidence of how library staff that participated in the iterative design process have already used what they have learned from the toolkit.

4.6: The project team will broadly disseminate information about both pathways of the DSTL through webinars and conference presentations. We will publish related results of focus group and testing research to leverage the iterative design process for usable insights for other projects in the field. All webinars will be free and publications will be available open access through a trusted institutional repository.

The DSTL planning grant involved librarians from any type of library who volunteered their time and developed a 40-person core design team. The implementation grant will continue the direct involvement of library staff in the **iterative design process** in multiple ways. First, it will formally involve representatives from 10 public libraries in a Core Design Team and these libraries will receive free Benchmark subscriptions. Second, we will offer incentives for participation of additional publicly invited focus group attendees who work in public libraries. Interactive design focus groups will include several modalities, including in-person and online (synchronous and asynchronous) opportunities. Third, as part of the evaluation process we will invite any interested library staff to review and test the initial designs and to provide feedback via a survey. Public webinars will be used to advertise the project and solicit interest in testing from a broader population of library staff, and ongoing opportunities for participation will be shared with the over 700 people who expressed interest in the planning grant. Library staff will be central to providing input for iteration and validation of final design. Iterative design will drive the development process through years 1 and 2. Feedback will be incorporated into the final design of both the create and explore pathways in the toolkit. An overall evaluation of the project's impacts will occur in year 3. Fully implemented, the DSTL will allow libraries to meet the real needs of their communities, including adding services and resources to support literacy for all.



*Library staff use interactive DSTL tools to create visualizations to support their story. They create a bar chart showing the gap between static collection investment and increased demand in book and e-book circulation. They use a line chart to show their own trends in Spanish-language programs and program attendance. They use community data to show how the library's offerings could better meet the demographic profile of the community. The resulting report incorporates both story and data visualizations and they share it with stakeholders. Library staff present a plan to increase literacy services for Spanish-speaking families at the library, with a special focus on children. They show how increasing the collection budget allocated to Spanish-language materials and increasing the number of Spanish-language children's programs would benefit more community members. The library director takes this data story—with visualizations and narrative based on **explore** and **create** pathways—to the city council budget presentation and to the library friends group to request a funding increase. Library staff will also share the data story with partners—including community organizations and the local school district—to get their input and support.*

Project Results

The DSTL will result in a two-pathway interactive toolkit that provides the guidance and evidence libraries need to tell powerful impact stories to communities and stakeholders. The long-term vision is cultivating data storytelling expertise as a signature expertise of our field, so that when communities have data storytelling needs, they are met at libraries and by librarians.

By design, the DSTL is highly adaptable. Making it generalizable and usable will require staying in dialogue with library staff as iterative design participants. The partnership between UIUC and PLA combines expertise in research, teaching, and professional development. Both organizations have rich histories of developing learning opportunities for information institutions and professional communities in the public library world and beyond.



Based on the evidence presented in the library's data story, the city council approves a modest budget increase, and the friends of the library group pledges to raise the additional funds needed. Spanish-speaking schoolteachers recommend new materials to purchase for the library's collections, and a local immigrant service organization volunteers to help run after-school library programs for children once a month. And the story continues, because the library staff now understand the importance of reaching audiences. Partnerships with community organizations make the most of these investments, through ongoing dialog about planning meaningful programming throughout the year. Library staff keep working, conducting in-person and online surveys of community response to these new investments. Adding Spanish language programs and connecting with community organizations leads to word-of-mouth publicity that increases both program attendance and foot traffic at the library.

Data is more available than ever before for public libraries, and yet there remains a gap in engaging with analysis and in training to communicate data effectively. With this implementation grant we aim to transform attitudes, practices, and collaboration among library staff in relation to data interpretation and evidence-based storytelling. The ultimate beneficiaries of this project are communities, because libraries are more effective when library staff understand and serve real, measurable community needs. Being able to better tell the story of their work will in turn help libraries secure funding and resources to continue to serve their communities into the future. Information must be conveyed to audiences that fund libraries so that they invest, intellectually and emotionally, in library improvement. The interactive toolkit will allow libraries to quickly start exploring their data, learning about storytelling best practices, and launching their own story creation. At this critical historical juncture, how libraries use evidence to serve their communities is more important than ever. The DSTL makes the connection between data and the stories that can have real impact.

Digital Products Plan

As outlined in the narrative, implementation of the Data Storytelling Toolkit for Librarians will include two pathways: **Explore** and **Create**. While these will be interlinked and cross-referenced, they will be developed using different technologies and infrastructure. Together they will provide greater opportunities for engagement with the DSTL across the library field and ensure its longevity.

In addition to the digital tools created through these pathways, PLA and UIUC will make the documentation, story templates, and any related publications freely and openly available to the public through a trusted institutional repository under a Creative Commons license.

Type

The **Explore** pathway will become part of an existing digital platform, Benchmark: Library Metrics and Trends (librarybenchmark.org). A project of the Association of College and Research Libraries (ACRL) and the Public Library Association (PLA), divisions of the American Library Association (ALA), Benchmark incorporates data from multiple sources and provides visualizations and interactive reporting tools that libraries can use to improve performance, demonstrate value, and increase impact in their communities. It launched in 2021 and ACRL and PLA are committed to maintaining and supporting it into the future. Benchmark is a custom-built platform hosted in a cloud environment and built on industry-leading tools (Microsoft Azure and the Microsoft stack, including PowerBI). The Explore pathway will involve the creation of a new “Stories” component within Benchmark, drawing on database infrastructure and functionalities that have already been developed. By expanding an existing tool, we take advantage of prior investments, infrastructure, and brand recognition, and we can ensure that the outputs of the grant project are maintained and updated over time.

The **Create** pathway will be a standalone website and the underlying toolkit used to deploy it. This will be composed of one reference deployment and implementation, as well as the underlying source code and software enabling others to deploy it elsewhere. This software will be licensed under a permissive, non-copyleft license (such as BSD 3-Clause or MIT) to enable reuse and integration into external systems. Deployment scripts and examples will be provided, as well as documentation (in markdown format) describing developing extensions to base functionality. Development of this toolkit, with the extension described above, will enable rapid iteration of new techniques, widespread deployment, and a more focused effort on storytelling, while still maintaining the ability of institutions and individuals to utilize their existing strengths and tools.

Availability

Explore: Benchmark is a web browser-based toolkit that does not require specialized software on the part of the user. Login is required for access; all US public libraries and state library administrative agencies have credentials for access (other types of users can request accounts). In their accounts, library users see dashboards – and will see “Stories” – pre-populated with their library’s data and customized peer comparisons. While there are limitations to availability because of the login required and types of access (described below), building this pathway in an existing tool that uses industry-standard infrastructure will help to ensure its longevity. It also increases ease of use for library users, since they do not need to prepare or format any data: it is all immediately at their fingertips in compelling and interactive visualizations.

Create: Like Benchmark, the dashboard toolkit developed through this project for independent deployment will be a browser-based toolkit with no additional requirements. As a result of its different focus, this will not require login for basic functionality, and the vast bulk of processing requirements will be client-side rather than requiring server-side computation. To the best of our ability, this will be built in a WAI-ARIA compliant manner and with accessibility-focused design choices.

Access

Explore: ALA holds copyright and intellectual property rights for Benchmark. For data that libraries submit through surveys on the platform, the library retains ownership of the data and grants ALA access to license, use, analyze, modify, distribute, and publish it. There are two levels of user access to the tool: 1) *free access* allows all libraries to participate in annual surveys and view resources and key metrics, and 2) a *subscription* offers upgraded access to a robust set of data dashboards and reporting tools. Subscriptions cost a maximum of \$400/year for public libraries, with discounts available. While the Explore pathway for this project will be developed within the subscriber portion of Benchmark, we will ensure that resources, documentation, and other publications are freely available. It is not feasible to make all components of Benchmark freely and openly accessible while also sustaining this tool, as outlined below.

Create: The purpose-built dashboard toolkit will not provide data natively, nor will it support server-side storage of data. All components of the source code will be made available under permissive licenses, as reference implementation will be freely accessible, and as a result of the focus on client-side computation we expect individuals to be able to use this as extensively as necessary. These are tradeoffs made in comparison with the significantly more functional Benchmark system.

Sustainability

Explore: Benchmark is not a one-time project output, but a robust and growing tool intended to support the library field long-term. PLA and ACRL update the data annually and make regular improvements to the functionality and user experience. The “Stories” component of Benchmark developed through this project will thus continue to be updated after the term of this grant, with new data added annually and technical capabilities updated as needed. While we would like to make all components of Benchmark free and open to all libraries, unfortunately that is not sustainable. Subscription revenue makes it possible for PLA and ACRL to continue to support and develop this tool and related research efforts. Currently Benchmark costs approximately \$150,000 per year in web hosting, support, and development expenses, not including staff time or other resources. Subscription revenue does not fully offset that cost as it stands; the tool is thus offered at below cost to libraries. PLA and ACRL have made a significant investment in this tool and are committed to ensuring it remains relevant and valuable to the field.

Create: Sustainability of open-source software is an open problem, but one that has a number of known best practices to mitigate risk. To that end, the software developed in this system will be strongly client-side focused (i.e., minimizing short- and long-term server demands), developed openly and with community-standard guidelines such as a set of contribution guidelines and a code of conduct, and we will endeavor to enable extensibility and flexibility of the underlying system to allow individuals to modify and provide changes upstream.

Data Management Plan

The goal of the DSTL is *not* to collect new data from public libraries or library staff, but to create tools that allow them to use their own or pre-existing data in new ways. Nonetheless, in the process this project will generate the following types of data:

- Software and its metadata,
- Community interactions and development history, and
- Educational materials, presentations, and publications regarding how to utilize, extend and adapt the toolkit.

Additional ephemera may be generated, but not retained beyond mandatory institutional policy requirements, such as website access logs.

Software

New software will be written both as extensions to existing platforms (explore pathway) and as new platforms in and of themselves (create pathway), as described in the Narrative and Digital Products Plan. End users will utilize the DSTL to generate data stories that will use their own or pre-existing data, and we will endeavor to ensure that the process of collecting, sharing, and preserving relevant metadata (and data) is as seamless and simple as possible for them. The software will not produce significant amounts of traditional data but will result in source code and related metadata and documentation.

The software for the **create** pathway will be developed in a completely open-source manner and released under a permissive Creative Commons license allowing re-use and re-distribution. Software will be openly accessible both prior to and during code review for acceptance into the appropriate main distributions. All code and documentation developed for DSTL will be held under a project organization on GitHub, with modifications to any individual projects as well as other utilized open-source projects contributed upstream in accordance with contribution guidelines. Prior to review and acceptance upstream the source will be stored either under a specialized account or under personal developer accounts, with the intention of merging upstream as quickly as possible. All end user-generated data and models produced by the developed software will be owned by the end user and under the end user's control. All software, deployment scripts, and other supporting tools will be archived in code repositories in use by the team's constituent members, unless precluded by existing licensing agreements. These include organization-level GitHub repositories and code repositories provided by the team members' universities and organizations.

Metadata for produced source code will be in the form of git changesets and code review comments, both of which are industry standard. If sample data is required for development, we will either generate it as a part of the project and license it under Creative Commons Zero or will solicit it from community members and distribute it only under their terms. Design documents created in the course of this proposal will be versioned and available freely both in original source form and in rendered form on the DSTL website or GitHub organization, and formatted in Markdown or other open formats as appropriate. Periodic releases of design documents and source code will be made and archived to a Zenodo repository, where they will be available in versioned and citable form. These releases will be mirrored to the appropriate institutional repository at the University of Illinois, either IDEALS or the Illinois Data Bank.

Implementing the Data Storytelling Toolkit for Librarians

The **explore** pathway will be built as a new component within PLA Benchmark. Benchmark is a custom-built platform hosted in a cloud environment using Microsoft Azure and the Microsoft stack, including PowerBI for data visualizations. For users the tool is browser-based, not requiring specialized software for access. ALA holds copyright and intellectual property rights for Benchmark. While Benchmark and its software are proprietary, for the DSTL we will capture procedural information and documentation related to the development of the explore pathway and make that openly and freely available to the public through a trusted institutional repository under a Creative Commons license.

Where we build on external software projects, we will submit our contributions upstream in a license-compatible way to enable their inclusion. Where extending existing systems, best efforts will be made to make extensions available under minimal restrictions and in compliance with software licenses.

Community interactions

During the iterative design and development process, we anticipate the production of code review comments, discussions (on mailing lists, in chat channels, etc.), user experience feedback, and other communication ephemera. This may include personally identifiable information (PII) where it is necessary to understand who the users are and follow up with them for additional discussion. However, any PII collected in the design process will be voluntary and not stored beyond the completion of this project.

Logs and other potentially sensitive information regarding access to websites deployed under this grant will be minimally retained, exclusively in compliance with institutional policies regarding privacy of participants. Logs of access to deployed websites, etc., will not be retained beyond times specified by institutional policies and will be stored in secure, privacy-compliant data storage facilities. Datasets will not be retained without explicit consent from users, and efforts will be made to rely upon client-side processing to minimize leakage of any personally identifying information.

Records of community code collections will be stored in similar systems, utilizing appropriate git changesets or through periodic archiving of software distributions. We will ensure that authorship of software and educational materials is correctly credited to those who participated in its development.

Educational materials

Documentation for DSTL components developed for this project will be version-controlled, reviewed by peer collaborators, and made publicly available. This may take the form of usage documentation, examples, recorded videos and coursework. Videos will be made available through YouTube or Vimeo channels with no restrictions on viewing. All documentation will be made available under Creative Commons licenses that enable reuse and modification. All documents, slides, papers, web material, and other forms of written data will be formatted using PDF, HTML, or Markdown, as appropriate. In this manner, we will endeavor to ensure that all educational materials are readily and freely available.

All documents, slides, papers, and other forms of data produced will be deposited into institutional repositories (Illinois Data Bank, etc) or made available through appropriate journals or online services (such as Arxiv.org, the Open Science Framework, FigShare, Zenodo, and SlideShare). Each of these works will be assigned a persistent Digital Object Identifier. Wherever possible, archives will be kept at multiple off-site locations, with backups utilizing both cloud providers (such as GitHub) and institutional providers.