**The New York Data Carpentries Library Consortium (NYDCLC)
Planning Grant**

ABSTRACT

The New York Data Carpentries Library Consortium (NYDCLC) comprised of The University of Rochester, Colgate University, Cornell University, and Syracuse University, and working in close collaboration with the regional library councils in central and western New York State, will plan, pilot, and assess a cross-institutional model to address two interwoven needs within contemporary librarianship. These are: (1) the need to increase librarians' capacity to meet their patrons' requirements for data and digital literacies that have become increasingly central to research, teaching, and civic participation; and (2) the need to collaborate across libraries and institutions to draw upon a range of expertise to meet those demands sustainably and equitably.

The pace at which technologies emerge, the many skills necessary to make use of these technologies, and the centrality of digital tools and information for library constituents' information needs means that advancing digital literacy is both a high priority for librarians and a daunting task. As demand for training increases, libraries must expand capacity to do this work in a sustainable manner. However, the growing needs are at odds with funding cuts in public libraries and primary, secondary, and higher education, which require librarians continually do more with less. As a result, the demand to work across disciplines, divisions, and institutions to generate solutions to large challenges within the library community is more urgent than ever.

During the course of the planning grant, the NYDCLC will focus on the following objectives and activities: (1) We will establish a framework for the consortium, including a steering committee and sub-committees tasked with specific areas of development and outreach. (2) We will partner with the regional councils to identify and recruit library staff who belong to groups underrepresented in public, secondary, and higher education librarianship to join the founding members on the steering committee. (3) We will conduct a regional needs assessment to help identify training priorities for skill-based workshops. (4) We will host three skill-based Carpentries workshops for librarians across the region. (5) Members of the steering committee will participate in The Carpentries instructor training. (6) We will host a one-day Symposium that will be free and open to the public. In the morning, we will pilot and assess concurrent hands-on workshops developed by the members of the consortium. In the afternoon, we will hold panel discussions where participants report on activities of the NYDCLC and lessons learned. (7) We will produce a white paper that will detail our structure and activities. The white paper will include rubrics, surveys, and data from our needs assessments and will include our meta-analysis of how the collaboration worked in practice, what we learned, and what we would change moving forward. This reflection on our work, structures, and processes will provide critical information for other groups seeking to foster similar regional collaborations.

While this planning grant focuses on training library professionals to address a high-priority gap caused by increasing demand for data support, the outcome of this work will include expanded access to information, skills, and networks of support for library patrons, whether in schools, public libraries, or institutions of higher education. The work of the NYDCLC will serve as a template for structuring a community of practice to support public and academic partnerships that meet the evolving needs of our profession.

**The New York Data Carpentries Library Consortium (NYDCLC)
Planning Grant**

STATEMENT OF BROAD NEED

The New York Data Carpentries Library Consortium (NYDCLC), comprising University of Rochester, Colgate University, Cornell University, and Syracuse University, seeks $82,688 in Laura Bush 21st Century Librarian Program Planning Grant funding in the Lifelong Learning category. The funding will support a 12-month effort to plan and pilot a collaborative skill-sharing program. The program will utilize The Carpentries model to expand the knowledge, skills, and abilities of librarians within central and western New York State to effectively support our diverse populations in the critical areas of data and digital literacies. This planning grant will provide the foundation for a three-year project that will address two interwoven needs within the library community: (1) the need to increase capacity to train and support librarians to meet the demands of data and digital literacies that have become increasingly central to research and teaching and the broader public's capacity for civic and cultural participation; and (2) the need to collaborate across libraries and institutions to draw upon a range of expertise to meet those demands sustainably and equitably.

Finding, using, and assessing information increasingly relies on a rapidly expanding set of digital tools and methods. In 2013, the American Library Association's Digital Literacy Task Force published a report on the need for libraries to be "on the front lines of digital inclusion and digital literacy efforts nationwide."[1] The authors define digital literacy as "the ability to use information and communication technologies to find, understand, evaluate, create, and communicate digital information, an ability that requires both cognitive and technical skills."[2] The report highlights some related issues that move beyond basic digital and technological competencies to include an understanding of the relationship between technologies, personal digital privacy, and how we might harness these tools for communication and collaboration. Further, the authors underline the fact that people need "these [digital literacy] skills to participate actively in civic society."[3] This latter point is also emphasized in discussions of digital equity in IMLS' report on its National Digital Infrastructures Initiatives. The IMLS report notes that people need access to technology, "digital literacy (including digital privacy)," and to possess the knowledge and tools "to use the technology effectively."[4] Indeed, to be digitally literate, one must be able to discover, evaluate, and create using a variety of digital tools and platforms. The pace at which technologies emerge, the skills necessary to support these technologies, and the centrality of digital tools and information for our constituents necessitates that finding a way to advance digital literacy is a high priority for librarians. This relentless but exciting pace also makes it a daunting task.

Digital tools and methods are continually evolving, so training on new approaches must be an ongoing commitment. Each technology requires deep expertise, which means that it is impossible for a single librarian to support the growing list of tools. Similarly, as demand for training and support for these tools increases, libraries must expand capacity to support this work in a sustainable manner. Finally, the growing needs are at odds with funding cuts in public libraries and primary, secondary, and higher education, which

---

[1] "Digital Literacy, Libraries, and Public Policy: Report of the Office for Information Technology Policy's Digital Library Task Force" (American Library Association, January 2013), https://www.districtdispatch.org/wp-content/uploads/2013/01/2012_OITP_digilitreport_1_22_13.pd, 1.

[2] Ibid., 2.

[3] Ibid., 2.

[4] Amy Rudersdorf et al., "National Digital Infrastructures and Initiatives: A Report on the 2017 National Digital Platform at Three Forum" (Washington, D.C.: Institute of Museum and Library Services, Office of Library Services, June 2018), https://www.imls.gov/publications/national-digital-infrastructures-and-initiatives-report-2017-national-digital-platform, 15.

means that librarians are continually seeking ways to do more with less. The need to work across disciplines, divisions, and institutions to generate solutions to large challenges within the library community is more urgent than ever. To address this need, library consortiums have formed around collection development, resource sharing, and contract negotiations with vendors, among other issues. For example, *Research Library Issues* (*RLI*) no. 296 focuses on the need for expertise across disciplines to tackle the challenges of Research Data Management (RDM).[5] While the issue of RDM is outside the purview of this proposal, its framing is centered on the need for human expertise and labor to come together to solve these issues. This is a concept that Nancy McGovern calls radical collaboration, and is particularly useful when thinking about the benefits of establishing communities of practice on the regional level to address the systemic issue of expanding capacity, sharing expertise, and supporting regional constituents. "The concept of radical collaboration means coming together across disparate, but engaged, domains in ways that are often unfamiliar or possibly uncomfortable to member organizations and individuals in order to identify and solve problems together, to achieve more together than we could separately."[6] Digital approaches, technologies, and literacies are advancing at such a rate that we need different perspectives and expertise in order to support them.

To meet the changing learning and information environment better, librarianship must include a stronger emphasis on digital and data literacies, which precipitates the need for continued education of librarians and information professionals in these critical areas. Specifically, there has been an increase in the last five years of international scholarship centered on the need for data literacy across all disciplines (see Literature Review in Supplementary Materials). This is, in part, because data generation has exploded, with 90% of the world's data generated in the past two years alone.[7] While data permeates every aspect of the scholarly, civic, and creative enterprise, the increase in mobile and smart technologies have made data more available, visible, and important to general audiences. Consequently, there is a broader need for individuals to understand how their personal data is being generated, collected, and used, which has led to public interest in data privacy. As questions of access and privacy build around public datasets, public libraries are being asked to host and make legible data for their communities. "But what often gets lost in the conversation is the idea of how public data should be collected, managed, and disseminated so that it serves everyone— rather than just a few residents—and so that people's privacy and data rights are protected. That's where librarians come in."[8] The widespread need for increased education and support around data requires cross-institutional support and solutions. The NYDCLC will provide *a scalable regional model* for how universities and public libraries can come together to share resources, expertise, and training for library professionals to meet the growing data and digital literacy needs of the public, as well as those of students and academic researchers.

The idea for creating the NYDCLC spawned from a conversation at a regional academic conference focused on data literacy: ENY/ACRL's 2018 Data, Libraries and Justice. A group of librarians from the aforementioned institutions in western and upstate New York gathered to address the challenge of providing adequate services for the increase in data-related projects and questions. Our campuses, like most of higher

---

[5] Judy Ruttenberg and Elizabeth Waraksa, eds., *Research Library Issues*, no. 296, 2018.

[6] Nancy McGovern, "Radical Collaboration and Research Data Management: An Introduction," *Research Library Issues*, no. 296 (2018): 6–22, https://doi.org/10.29242/rli.296.2, 6.

[7] Bernard Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes Now*, May 21, 2018 https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#127565ef60ba; Gary Kim, "90% of All Data Generated in the Last 2 Years," *IP Carrier: A Chronicle of Business Model Change and End User Transformation in the Global Communications Industry*. (blog), June 2, 2017, https://ipcarrier.blogspot.com/2017/06/90-of-all-data-generated-in-last-2-years.html.

[8] Linda Poon, "Should Libraries Be the Keepers of Their Cities' Public Data?," *CityLab*, February 11, 2019, https://www.citylab.com/life/2019/02/libraries-public-information-city-data-digital-archive/581905/.

education, face a dearth of support for data, especially given its recent ubiquity. We agreed that librarians, not only on academic campuses but also in surrounding communities, are uniquely positioned to teach data literacy skills given their expertise in finding, accessing, and evaluating interdisciplinary information. We recognized the need for collaboration and training at our institutions in order to provide the necessary support for reading, comprehending, using, and expanding upon the data users encounter. The planning grant will facilitate and formalize a framework and a set of initial activities of the consortium. It will allow us to conduct local and regional needs analyses, pilot and assess librarian training models, and develop strategies for building mutually beneficial partnerships with our regional library councils to expand the consortium to include public and secondary school librarians. The planning year is crucial for developing a model that will be sustainable and allow us to scale up the project in future years to have the greatest and broadest impact on library professionals across our region. Furthermore, it will provide a sound framework to the broader library community that other regions may use as a template for how to do this work.

NYDCLC has identified The Carpentries as a sustainable pedagogical model for supporting our own learning and for sharing new skills with academic and public library staff. The Carpentries is a not- for-profit network of volunteers, comprised of more than 70 member organizations in 10 countries that teach foundational coding and data science skills to researchers worldwide. The Carpentries' mission, "to build global capacity in essential data and computational skills for conducting efficient, open, and reproducible research"[9] aligns with our needs for capacity building within data and digital literacies. Membership in The Carpentries will enable the NYDCLC to develop further critical skills, including effective pedagogical approaches to teaching digital literacies. Members receive priority access to instructor training and guidance with capacity building at their organizations. Community-developed lessons enable members to share globally the effort for developing impactful lessons on technical topics. While The Carpentries provides a model and a means for training librarians, which may have been sufficient for our individual institutions, the NYDCLC acknowledges the broader need of engaging with our communities in order to tackle these issues at a regional level through *a community of practice*. As such, we are committed to developing curricula and infrastructure to support librarians and staff in both public and scholarly spheres, through relationships developed between academic institutions as well as with regional and national library councils. This broader regional structure will allow NYDCLC to support the data needs of faculty, students, and the broader public within their diverse communities. These councils already include Central New York Library Resources Council (CLRC), the South Central Regional Library Council (SCRLC), Rochester Regional Library Council (RRLC), and The Delaware-Chenango-Madison-Otsego Board of Cooperative Educational Services (DCMO-BOCES). Throughout this year in planning, the NYDCLC will design infrastructure to meet the needs of our local institutions and regional organizations as well as develop a strategic plan for how to sustain and expand this model so that it can be replicated by other regions and other communities of practice.

While the need for data literacy training is evident from both pedagogical scholarship[10] and our members' first-hand experiences, the IMLS planning grant will allow for a comprehensive needs analysis across the regional library councils to identify and plan workshops tailored to specific needs. Although a coordinated environmental scan between academic and public libraries of the central and western New York region has never been conducted, interest and need is evidenced through requests for professional development opportunities and workshop participant feedback. For example, through previous workshop participant feedback, Rochester Regional Library Council reports 20 requests for continuing education in

---

[9] "The Carpentries Handbook," The Carpentries, accessed March 8, 2019, https://docs.carpentries.org/topic_folders/governance/bylaws.html#purpose.

[10] Tibor Koltay, "Accepted and Emerging Roles of Academic Libraries," *Supporting Research 2.0. The Journal of Academic Librarianship*, 45(2): 75-80. https://doi.org/10.1016/j.acalib.2019.01.001.
L. Lyon, et al., "Applying translational principles to data science curriculum development," *Proceedings from iPress 2015: 12th International Conference on Digital Preservation*. Retrieved from: https://phaidra.univie.ac.at/o:429552.

digital humanities and technology from September 2018 through February 2019. At the same time, the University of Rochester saw an increasing demand for data-related professional development and outside-of classroom learning for students, staff, and faculty. From September 2018 to February 2019, there were over 40 unique participants in data related training such as Introduction to R, Tableau Basics and Data Management workshops. One of these was a Tableau workshop requested by and run for librarians that drew over 15 participants. Graduate students at the University of Rochester have expressed interest in learning more about data through events that they have organized, including one that focused on humanistic approaches to data visualization and one that engaged experts at local non-profit organizations in discussions of data and social justice. Librarians at Syracuse University have also seen increasing demand for data training, especially related to undergraduate research. They have developed programming and invited external experts to their *Issues in Digital Scholarship* series to address data literacy issues for faculty and graduate students. At Cornell, librarians currently provide workshops and individual consultations in data acquisition, cleaning, analysis, and visualization for researchers in natural sciences, social sciences, and the humanities, but struggle to keep up with increasing demand for this type of research support. While these institutional examples provide a brief snapshot of our current efforts and challenges, they all speak to a broader need to increase our capacity for data support and trainings. Moreover, our individual institutions lack the capacity to organize the data and digital literacy offerings into a reliable, comprehensive, and pedagogically rigorous whole. While each institution has specific expertise and the ability to support a limited number of faculty and students, the general knowledgebase among our own librarians is lacking. The NYDCLC is positioned to meet these needs through collective expertise and our commitment to developing robust and reproducible curriculum that will be shared with our region and the broader library community.

In addition to establishing the specific data needs of the private and public libraries in central and western New York State, this planning grant will provide a unique opportunity to support library staff as facilitators of lifelong learning and digital literacies in an area challenged by both rural and urban demographics (see Diversity Plan). The Carpentries train-the-trainer model provides both skill-focused and pedagogy-focused training, e.g., librarians will develop technological competencies, and learn effective ways to teach those competencies to others. Thus, the NYDCLC will provide education and training programs to increase the skills and expertise of librarians who will in turn, be able to facilitate their patrons' learning of data literacy skills. The NYDCLC will expand the train-the-trainer model to integrate librarians from the surrounding communities and foster timely collaborations to address issues of shared concern around patron support and learning. Community-based events, like Colgate University Libraries' 2018 Endangered Data Week, have proven that these collaborations can be successful. Hosted in the community, Colgate librarians partnered with public librarians to facilitate a shared conversation about county and state data collection and use. This event drew over 50 attendees from a diverse cross-section of the Madison County community.[11] The NYDCLC will enable librarians to learn new skills in a rapidly changing environment as well as how to integrate digital pedagogy into their information literacy instruction. By broadening the scope beyond academic librarianship to include public, school librarians, and information professionals, the NYDCLC expands its impact across a larger and more diverse patron base. Unlike MOOCs or other online tutorials, these in-person inquiry-based, cross-disciplinary workshops ensure that all individuals and communities have the information technology capacity needed for full participation in our society, democracy, and economy through hands-on training and personalized workshops.

PROJECT DESIGN

Conversations about the structure of the consortium, the design and distribution of internal and external environmental scans, and the development of potential Carpentries workshops have already begun and will

---

[11] Sarah Anderson, "Colgate Joins Nationwide Endangered Data Week with Environmental Justice Panel," *The Colgate-Maroon News*, March 8, 2018, accessed 3/1/2019, http://www.thecolgatemaroonnews.com/news/article_c0fe25a6-22ea-11e8-9cca-57c9ddae4f7e.html.

continue up to the starting date of the planning grant in September 2019. Through this pre-planning process, the initial eight members of the steering committee (see List of Key Members) will help identify and organize potential workshops around the region and develop structures to ensure a smooth onboarding process for the newly identified librarians joining the steering committee from the regional councils. Below we describe a phased approach that will enable sufficient time for planning, executing, and assessing the activities of the planning grant. As reflected in the Project Timeline, due to the academic calendar, holiday breaks, and winter weather in central and western New York, during December through February we will minimize travel and consortium workshops. During this time, virtual meetings and workshop planning will continue between the steering committee and regional councils.

The steering committee will work to mitigate a few potential risks that have been identified during the initial phase. First, there may be unanticipated changes in data needs over the long term, and the consortium will need to be flexible enough to shift or adapt to address them. The pedagogical aspects of the training and the workshop design skills will help the committee to prepare for this by training librarians in the best practices for teaching skill-based workshops, regardless of the technology. As new trends and tools emerge, we will be able to develop new workshops based on The Carpentries model to meet the changing needs of our audiences.[12] We chose The Carpentries as our training model because of its well-established record of accomplishment in successful extra-classroom pedagogy for digital skills. With 76 member organizations worldwide, The Carpentries has trained over 1,600 instructors and affected 38,000 learners through workshops in 46 countries. The challenge of sharing skills and expertise across institutions presents an additional risk, as does the challenge of recruiting a set of diverse institutional partners. The steering committee will seek to recruit, communicate, train, and support with all potential risks in mind, and intends to design mechanisms to address/mitigate them.

Phase 1: Plan and Attend Coordinated Carpentries Workshops across the Region
Beginning in the early summer 2019, the steering committee will formalize and assign roles among its members. The steering committee will outline requirements for other institutions' participation in the consortium and schedule three initial data workshops for September and October aimed at an audience of regional librarians and archivists. At least one of these workshops will be attended by all eight core members of the steering committee and will be held at a regional council space, public library, community college, or one of the State University of New York campuses. To ensure geographic distribution and increase participation, workshops will be held near each of the four partner institutions. The workshops will be free for attendees and marketed to librarians and archivists via our institutional channels and through partnerships with the regional library councils. To further encourage participation and extend access, the NYDCLC will provide five bursaries for each of the skill-based workshops—for a total of 15 bursaries at $400 each—to help cover hotel and per diem costs for librarians who might otherwise not be able to attend.

These workshops will cover a broad range of topics from the three main areas of focus for The Carpentries' lessons: software, data, and library. For example, the Library Carpentries Workshop introduces terms, phrases, and concepts in software development and data science, how to best work with data structures, and use regular expressions in searches. This workshop introduces the Unix-style command line interface, and teaches basic shell navigation, as well as the use of loops and pipes for linking shell commands. It also introduces global regular expression (grep) tools for searching and subsetting data across files. Exercises cover the counting and mining of data. In addition, the workshop covers working with OpenRefine to transform and clean data, and the benefits of working collaboratively via Git/GitHub using version control to track work.

Moreover, a Data Carpentries workshop teaches the core skills for working with data effectively and reproducibly. When working with data, it is often difficult to figure out what skills to learn or where to start.

---

[12] The Carpentries, "The Carpentries 2018 Annual Report," accessed 3/1/2019, https://carpentries.org/files/assessment/TheCarpentries2018AnnualReport.pdf, 5.

The focus of this workshop will be on identifying the fundamental skills needed in a given domain and developing/teaching these skills in hands-on, two-day, interactive workshops. Workshops are designed for librarians with little to no prior computational experience and are domain-specific, so that librarians are working with data most relevant to their own disciplines. The lessons follow a narrative structure, working with one dataset through the whole research lifecycle from data and project organization to analysis and visualization.

Collectively, these initial, coordinated workshops led by certified Carpentries instructors, will allow the steering committee an opportunity to gain content knowledge, learn in a collaborative environment with other regional librarians and archivists, generate interest in the NYDCLC, and gather feedback from participants on the workshops as well as solicit requests for future topics and lessons. Because each workshop will be open to approximately 40 participants, we hope to reach at least 120 librarians and archivists from a cross-section of public, academic, special, and school libraries in the region.

Phase 2: Expanding the Steering Committee and Conducting Environmental Scans

During the initial workshops, and in collaboration with the regional councils, we will identify and recruit underrepresented minority librarians from public and higher education institutions across central and western New York State to join the steering committee. Members will be selected based on a combination of nominations, attendance at coordinated workshops, self-interest, and strategic recruitment. This process will utilize pre-existing application forms created by the regional councils to award professional development grants. The seven additional steering committee members will be required to be employed by an academic, school, or public library within the geographic purview of the regional library associations and commit to leading at least one Carpentries workshop for their institution or regional library association after becoming a certified Carpentries instructor.

Together, the fifteen-person steering committee will create the organizational structure of the consortium through subcommittees focused on specific tasks and activities. For example, an initial working group will be formed to perform an environmental scan to investigate and define our different institutional (Rochester, Cornell, Syracuse, and Colgate) approaches to and need for data literacy and pedagogy. This initial report will include a list of all library instruction courses or workshops in the last five years with a data literacy component, a survey of librarians and staff on perception of data literacy tools (R, Python, Figshare, etc.), and confidence in working with data. This environmental scan will help us better understand what data resources and support programs already exist on each campus, determine which librarians and departments are engaged in data literacy, and identify consortium strengths and gaps. A second working group will collaborate with the regional library councils to perform a community needs assessment across individual regions to identify pre-existing professional development opportunities, data literacy needs, and challenges to participation. This regional report will list all workshops and webinars held in the last five years related to data literacy (complete with attendance data and assessment) and a similar survey of librarians and staff on perception of data literacy tools and confidence in working with data.

Together, the institutional and community environmental scans will serve as the basis for a white paper that not only summarizes the need of academic and public librarians in the regional consortium, but also provides a baseline assessment of data literacy needs and opportunities among librarians within The Carpentries model. As reported in the 2018 *Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys*, though many of The Carpentries' workshops are hosted on university or college campuses and other research-based communities the majority of attendees are graduate students (50% or 5,009) compared to librarians/archivists (225 or 7%). Moreover, 66% of all attendees, across all fields, are early career professionals.[13] These environmental scans will assist the consortium in cultivating and

---

[13] Kari Jordan, François Michonneau, and Belinda Weaver, "Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys" (The Carpentries, July 17, 2018), Zenodo, http://doi.org/10.5281/zenodo.1325464, 18.

sustaining regional partnerships and inform curriculum development for region-specific Carpentries workshops. It will also provide the broader library community with the rubrics and surveys developed out of this work, so that other regions interested in performing similar gap analyses may use and adapt our model more readily.

Phase 3: Carpentries Pedagogy Training

In late fall/early winter 2019, the fifteen steering committee members (eight from List of Key Members and seven identified during Phase 2 of the planning grant) will undertake The Carpentries instructor training together, in preparation to lead local workshops during the spring and early summer. The community needs assessment performed in advance of the training will provide guidance in determining how many workshops to conduct and in which parts of the region they should take place. Steering committee members who participate in the Data Carpentries pedagogy training will each agree to teach at least one workshop within their region within a year of the training.

The initial eight steering members from partner institutions will bring expertise from across the spectrum of digital scholarship. For example, Emily Sherwood, Eliza Bettinger, and Jesi Buell all have expertise in developing workshops to help train librarians in digital scholarship methods and tools, as well as design-thinking pedagogy. Similarly, Patrick Williams teaches IST 600: Digital Humanities for Librarians, Archivists, and Cultural Heritage Workers at the Syracuse University School of Information Studies, a course he developed to prepare current and future information professionals to engage digital literacies in positions across the wide range of GLAM settings. Adrienne Canino, Zoë Wilkinson Saldaña, and Joshua Finnell have all led cross-departmental initiatives in data literacy and data outreach services. Finally, Déirdre Joyce, former Assistant Director at the Central NY Library Resources Council (CLRC), has lead metadata initiatives and projects at Syracuse University.

The additional seven steering committee members, identified through the regional councils, will enhance both the breadth and scope of digital literacy training through their own areas of expertise and disciplinary training. In particular, by recruiting a diverse cohort—institutionally and demographically—the steering committee will comprise a multitude of pedagogical perspectives and subject-specific expertise. More importantly, a shared conversation around teaching and learning will emerge through our collective experience of The Carpentries instructor training. The Carpentries Instructor Certification workshop consists of two-days of short lessons that alternate between individual and group practical exercises, including practice teaching sessions, followed by a series of follow-up assignments. In total, this planning grant will create 15 Carpentries instructors certified to teach Software Carpentry, Data Carpentry, and Library Carpentry workshops. Equally important, all instructors will adhere to The Carpentries Code of Conduct, also adopted by the NYDCLC, dedicated to providing a welcoming and supportive environment for all people regardless of background or identity (see the Diversity Plan for more information about the code of conduct).

Phase 4: Symposium and Assessment

In the summer of 2020, the steering committee will host a Symposium at one of the regional campuses in order to draw a community of librarians and archivists together, along with university librarians, directors, and other stakeholders. The Symposium will include a series of hands-on workshops in the morning that will provide an opportunity for instructors to pilot and assess the workshop materials generated during the planning year. The afternoon will include panels and roundtables where participants report on activities of the NYDCLC and lessons learned. The consortium will work in conjunction with the regional councils to handle registration forms, advertise through listservs, and coordinate venue logistics. Like the skill-based workshops, the Symposium will be free for participants and the NYDCLC will provide ten $400 bursaries to help cover travel costs for librarians who would otherwise be unable to attend. This one-day Symposium will have the following outcomes:

- Provide an opportunity for outreach to share the results of the institutional and regional environmental scans with the community, as well as the lessons learned over the course of the planning grant.
- Create an open dialogue around shared strengths, perceived weaknesses, and strategic directions for the consortium moving forward.
- Pilot developed workshop materials. The 15 members of the steering committee will collaborate to lead concurrent workshops for the attendees, showcasing both their pedagogical skills and demonstrating the train-the-trainer model to a wider community.
- Perform an initial assessment on workshops. In order to understand the impact of our instructor-training program, we will collect information about attendees' attitudes and skills related to our content before and after each workshops utilizing The Carpentries feedback assessment form.
- Steering committee members will meet with representatives from the regional councils to plan and lead at least one Carpentries workshop in the 6 months following the symposium.

The presentations, workshop materials, and feedback gathered at the symposium will inform the submission of a future project grant proposal and be made publicly available. All slides presentations, survey data, promotional materials, posters, and white papers will be deposited in the NYDCLC GitHub repository (See Digital Product Form).

DIVERSITY PLAN

By working through the regional councils, we will identify and recruit the initial steering committee library staff from across central New York from among those belonging to groups underrepresented in public, secondary, and higher education librarianship. To minimize financial barriers to participation, the planning grant will provide travel stipends for meetings, workshops, Carpentries instructor training, and attendance at the Symposium. These costs represents the vast majority of our proposed budget to IMLS. Rochester, Cornell, Syracuse, and Colgate have an aggregated endowment of 12.1 billion dollars and as outlined and acknowledged in the organizational profile, all four academic institutions serve a student body with less than 50% students of color. However, the communities surrounding each of these universities is both racially diverse and financially disadvantaged. For example, the two major urban areas in the region, Monroe County (Rochester) and Onondaga County (Syracuse), experience the fifth and seventh highest racial income disparities among older industrial communities in the U.S., according to the 2018 Brookings Institution's Metropolitan Policy Program report. Syracuse is among the ten poorest U.S. cities with populations over 65,000, according to Census Bureau data and Rochester is ranked 12th.

Angela Siefer, Director of the National Digital Inclusion Alliance, astutely points out that, "The goal of digital equity is achieved through digital inclusion, when users have access to technology (such as WiFi), possess sufficient digital literacy to use that technology, and have the tools (e.g., software and hardware) to use the technology effectively. Also, users need to feel safe employing the technology, understanding, for example, the privacy risks related to its use."[14] Many of communities in central and western New York rely upon regional school and public libraries to provide a safe and welcoming place to develop and obtain digital and data literacy skills. As a corollary, it is important that the steering committee of the NYDCLC is representative of the communities it aims to serve. Having representation from and engagement with library workers from diverse and underrepresented groups—including but not limited to race, ethnicity, gender identity or expression, sexual orientation, age, ability, nationality, religion, skill sets, and experiences—will position us to address the changing needs of our members and their communities. Moreover, by working in collaboration with the regional councils to create a mutually beneficial relationship, advertising Carpentries

---

[14] Angela Siefer, "What Do We Mean When We Say 'Digital Equity' and 'Digital Inclusion'?," *Benton Foundation* (blog), submitted October 27, 2016, https://www.benton.org/blog/what-do-we-mean-when-we-say-digital-equity-and-digital-inclusion.

workshops across school and public libraries will ensure that the curriculum is reaching a broad audience of participants beyond the university walls of Rochester, Syracuse, Cornell, and Colgate.

Furthermore, the NYDCLC will develop and adhere to a code of conduct statement that articulates an intentional commitment to diversity, equity, and inclusion in all aspects of our consortium. Everyone who participates in Carpentries activities is required to conform to this Code of Conduct. It applies to all spaces managed by NYDCLC including, but not limited to, meetings, workshops, email lists, and online forums such as GitHub, Slack and Twitter. Workshop hosts are expected to assist with the enforcement of the Code of Conduct. To ensure that workshops are accessible and inclusive, all workshops locations and meeting rooms will be ADA-compliant and all instructional and workshop materials generated by the consortium will adhere to the Web Content Accessibility Guidelines established by the W3C.

BROAD IMPACT

The goals of the NYDCLC planning grant align with the IMLS mission as noted in the Notice of Funding Opportunity for FY2019 for the Laura Bush 21st Century Librarian Program, "to inspire libraries and museums to advance innovation, lifelong learning, and cultural and civic engagement."[15] Through this grant, the NYDCLC will pilot, assess, and provide a roadmap for developing a community of practice that draws on the expertise of its members in order to expand access to training that will increase facility with digital and data literacies for librarians. The train-the-trainer model described in our Project Design will not only build capacity and the knowledge base of librarians across the region, it will allow the librarians, in turn, to more fully support the learning and literacy of their patrons through innovative workshops that will be developed, assessed, and shared through the planning grant.

While this planning grant in the Lifelong Learning category focuses on training library professionals to address a high priority gap caused by the increase in demand for support for data, the outcome of this work will include expanded access to information, skills, and networks of support for library patrons, whether in schools, public libraries, or institutions of higher education. The twenty-four counties served by these regional councils and universities include 2.8 million people, primarily located in rural and underserved urban areas. The education levels for people over the age of 25 varies: 10% did not graduate from high school, 30% finished high school, 30% attended some college or received an associate's degree, and 17% received a bachelor's degree. The median household income is below both the state and national median in all but two counties in the region. Further, over 14.8% of people in the region were living in poverty at some point in the last year, a higher percentage than both the state and national averages. The increased capacity to support the changing information needs of these populations has the potential for both broad and significant impact.

Other academic institutions have sought ways to increase support for emerging digital practices through cost-sharing training programs, like the New England Software Carpentry Library Consortium (NESCLC); however, they have yet to address the larger needs of the public or provide a model that facilitates robust collaboration across their institutions. As Matt Burton et al. discuss in *Shifting to Data Savvy: The Future of Data Science in Libraries*, a particular concern and challenge for the library community in sustaining data literacy initiatives is the development of both a structure and a diverse group (public and academic libraries) of stakeholders.[16] Barriers to sustaining a model are varied yet common to almost all library initiatives: the silo effect, scale, resources, credibility, and culture. By expanding beyond a consortium of academic institutions to include regional library councils, we intend to model a public-academic partnership that is designed to overcome these barriers.

---

[15] "Laura Bush 21st Century Librarian Program: FY2019 Notice of Funding Opportunity" (Institute of Museum and Library Services, 2018), https://www.imls.gov/sites/default/files/fy19-ols-lb21-nofo.pdf, 3.

[16] Matt Burton et al., "Shifting to Data Savvy: The Future of Data Science in Libraries," Project Report (Pittsburgh, PA: School of Computing and Information, University of Pittsburgh, 2018.), http://d-scholarship.pitt.edu/id/eprint/33891, 14.

In fact, we believe that these public-academic collaborations are necessary to leverage institutional expertise and funding to increase broad impact and address some of the larger issues facing libraries. We have seen successful models of public-academic partnerships that address separate, but related, issues around data preservation and access, such as the 2017 IMLS National Leadership Grant awarded to The University Library System at the University of Pittsburgh. This successful project was established in partnership with the Carnegie Library of Pittsburgh, the Western Pennsylvania Regional Data Center, and the Urban Institute to develop a regional platform for local, open, civic data. The consortium model underscores the importance of building infrastructure and workflows to provide access—in this instance to data about one's city, county, or state—but also serve to address an additional need to facilitate the acquisition of skills necessary to understand and use the collections by leveraging critical stakeholders.

This planning grant will have a significant and broad impact on local, regional, and potentially national librarians to foster the critical skills of discovery and cultivate critical thinking. It relies heavily on planning for successful collaboration between four institutions as well as across communities and organizations in central and western New York. Professional development for librarians in the area of data literacy and digital technology is a lifelong learning initiative due to its ever-evolving nature. By learning these skills and training others, the workshops will enable hands-on, self-directed learning, and substantial opportunities for improved understanding and increased access for the diverse communities in upstate New York. At its heart, NYDCLC is driven by core library principles like access, lifelong learning, democracy, and social responsibility. An initial planning grant will help define and scope a sustainable consortium model through piloting this first year of training and organizational development. The NYDCLC will facilitate a mutually beneficial and cross-institutional model for developing and implementing inquiry-based programming to meet the needs of students and researchers in academic and public domains. The consortium will design a workflow for how to facilitate a more inclusive, democratic, and diverse partnership and recruitment model as it expands across the region.

Ultimately, the work done by the NYDCLC will be a template for how to structure a community of practice to support public and academic partnerships to meet the needs of our evolving profession. At the end of our planning grant, the NYDCLC will produce a white paper that will detail our structure and activities. As noted in the Project Design, the white paper will include rubrics and surveys related to our institutional and regional needs assessments that could be readily adapted. Further, the anonymized data from our needs assessments could be used as the basis for a data driven argument for this type of initiative and act as a benchmark for assessing needs in other communities. Perhaps the most critical part of this report will be our meta-analysis of how the collaboration worked in practice, what we learned, and what we would change moving forward. This reflection on our work, structures, and processes will provide critical information for other groups seeking to foster similar regional collaborations. To ensure that our reproducible model is publicized broadly to both academic and public libraries, the steering committee members will submit panel and/or poster sessions to both the 2021 American Library Association in Chicago and the 2021 Association of College and Research Libraries in Seattle. Our white paper will be distributed to our professional organizations and shared with regional council networks. Further, to reach a broader public audience, the steering committee members will commit to hosting a minimum of eight webinars for State Library Commissions across the country. For example, we may share our work as part of NCompass Live, the Nebraska Library Commission's weekly webinar series. At the conclusion of this planning grant, we will have developed both a template and roadmap for public-academic library partnerships across the country to meet challenges facing librarians in developing and maintaining ever-evolving data literacy skills.

SCHEDULE OF COMPLETION

Activities	2019				2020								
	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept
Steering Committee Meetings	████	████	████	████	████	████	████	████	████	████	████	████	████
Institution(s) Environmental Scan	████												
Steering Committee & Regional Council Meetings		████	████	████	████	████	████	████	████	████	████	████	████
Regional Councils and Steering Committee Gap Analysis and Environmental Scan		████	████										
Nomination of librarians from regional councils		████	████										
Host 3 Coordinated Carpentries workshops		████	████										
Carpentries Pedagogy Training & Certification							████	████	████				
Develop NYS Data Carpentries Consortium Structure									████	████			
Assessment/Symposium Planning										████	████	████	
Symposium												████	
Final Report and Dissemination of Work													████

DIGITAL PRODUCT FORM

Introduction

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

Instructions

All applications must include a Digital Product Form.

☐   Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

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

Part I: Intellectual Property Rights and Permissions

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

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

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

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

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

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

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

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

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

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

C. Metadata

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

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

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

D. Access and Use

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

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

Part III. Projects Developing Software

A. General Information

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

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

B. Technical Information

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

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

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

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

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

C. Access and Use

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

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

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

Name of publicly accessible source code repository:

URL:

Part IV: Projects Creating Datasets

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

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

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

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

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

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

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

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

Name of repository:

URL:

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