Expanding Library Carpentry

Abstract

Librarians have an increasingly important role to play in mediating access to and use of the critical information technologies and resources integral to 21st century culture, commerce, and scholarship. In order to fulfill this role successfully, librarians must take advantage of training and professional development opportunities to remain current with the rapidly evolving information landscape, particularly with regard to computational and data-driven methods. A nascent Library Carpentry (LC) movement has arisen internationally that seeks to emulate the success of the Data Carpentry (DC) and Software Carpentry (SWC) initiatives in providing librarians with the critical computational and data skills they need to serve their stakeholders and user communities. California Digital Library (CDL) is requesting $249,553 in IMLS Project Grant funds in the Curating Collections category for Continuing Education to further advance the scope, adoption, and impact of Library Carpentry in the US.

Our 2-year project focuses on these main strands of activity: (1) development of a core curriculum of training modules based on DC- and SWC-authored material but optimized for librarian instruction; (2) regionally-organized training opportunities for librarians, leading to an expanding cohort of certified instructors available for librarian training in critical skills and tools, such as command line shells, OpenRefine, Python, R, SQL, and research data management (RDM); and (3) community outreach to raise awareness of Library Carpentry and promote the development of a broad, engaged community of support to sustain and advance the movement and its impact.

The training opportunities enabled by our project will provide librarians with the critical skills and tools they need to participate meaningfully in the development and use of an evolving National Digital Platform for library content and services, and to be effective digital stewards in Curating Collections. Today's academic library should be at the forefront of computational and data-driven pedagogy. Through our efforts to substantively empower librarians with vital computational and data literacy, this project will allow academic libraries to thrive as centers for 21st-century scholarship and learning, anchors for engaged and informed scholarly and public discourse, and stewards for effective digital custodianship.

Expanding Library Carpentry in the US

While information needs have increasingly come to depend upon sophisticated usage of data and software, professional development for many communities, including librarians, has not kept pace. In response, workshop-based approaches to learning have found success, especially with the well-established, self-sustaining Data Carpentry (DC) and Software Carpentry (SWC) "train-the-trainer" programs.[1] However, these programs offer hurdles for the library community in terms of scope, level of prior data literacy experience, and familiarity with computational concepts. In response, a parallel Library Carpentry (LC) initiative has begun as an international grassroots effort to provide software and data management training customized for librarians.[2]

This project will catalyze the success of the LC community efforts and facilitate the transition of LC to a sustainable skills training component of IMLS's national digital platform. To execute on this goal, California Digital Library (CDL) seeks $249,553 in Laura Bush 21st Century Librarian Program Project Grant funding in the Curating Collections category to support a 2 year effort to promote and coordinate Library Carpentry activities for continuing education of US librarians.

1 Statement of Need

In today's highly interconnected, online, and data-driven world, librarians play a key role in supporting the diverse information needs of the scholarly enterprise. As institutions have begun to view the data output of researchers as an asset, and as requirements for data sharing have become the norm for funding, issues around research data best practices have emerged. This has resulted in the formation of Research Data Management and Research Data Curation programs, often located in the library.[3] The services offered include access to repositories, information on Data Management Plans (DMPs), and consultations about requirements. While studies show this kind of training is a critical need for researchers, training in data skills and tools are lacking[4], and librarians with the ability to teach or help researchers with their hands-on data needs are rare.

---

[1] http://www.datacarpentry.org/; https://software-carpentry.org/

[2] J. Baker et al. (2016), "Library carpentry: Software skills training for library professionals," *LIBER Quarterly* 26(3):141–162, http://doi.org/10.18352/lq.10176; see also https://librarycarpentry.github.io/

[3] Flores, J. R., et al. (2015). Libraries and the research data management landscape. In J. C. Maclachlan, E. A. Waraksa, & C. Williford (Eds.). The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy (pp. 82–102). www.clir.org/pubs/reports/pub167/pub167.pdf

[4] Federer, L. M., Lu, Y. L., Joubert, D. J., Welsh, J., & Brandys, B. (2015). Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. PLOS ONE, 10(6), e0129506. https://doi.org/10.1371/journal.pone.0129506

In addition, librarians cannot be effective in connecting scholars with information and tools without themselves understanding how research is done, the nature of scholarly communications, and the tools used. In the past, the relationship between scholars, faculty researchers and librarians was based on a service model: the 'user' would approach the librarian for information and receive it. Today, that relationship must evolve into a collaborative model, as 'users' now need information from their own or other data sets, and come to the library for help in analyzing and/or extracting it.

This means librarians need to become skilled in data best practices, including data management and analysis during their formal educational training. This also means that existing librarians require ongoing professional development to stay up to date on the evolving landscape of skills and tools, such as OpenRefine, Python, R, and SQL, and data-intensive approaches to librarianship.

This project aims to enhance and multiply the opportunities for librarian training in critical data and software-related skills, and by doing so, help to integrate librarians into the larger communities of research and learning at their institutions. LC provides the best means for achieving these goals at a national scale.

2 What is Library Carpentry?

Library Carpentry leverages the success of the Data Carpentry and Software Carpentry pedagogy, which is based on providing a goal-oriented, hands-on, trial-and-error approach to learning[5]. In this model, "instructors" are certified in the Carpentry way of teaching and are asked to commit to building new lessons, reworking/improving existing lessons, and offering workshops in their regions. Each workshop is hands-on where attendees go through lessons with their certified instructor. The goal is to become practitioners while in the workshop and then offer continuous support through online and in-person community interaction afterward.

With their "train-the-trainer" model, both DC and SWC are both built to create learning networks and find success in creating active communities. They have used this model to scale with parallel tracks of developing lessons, offering workshops, and expanding the community.

The nascent LC movement began as a way to build a data and software savvy librarian community while also offering a low-barrier way for librarians to join efforts with the DC/SWC community. Not originally a formal partnership with DC/SWC, LC began as a loose translation of their approach to training. It was immediately successful in attracting attention and scaling. However, LC workshops were not originally conducted

[5] T. Teal et al. (2015), "Data carpentry: Workshops to increase data literacy for researchers," *International Journal of Digital Curation* 10(1): 135-143, http://dx.doi.org/10.2218/ijdc.v10i1.351

by certified DC/SWC instructors and the quality of approach and sustainability was immediately questioned.

Recent efforts to bring LC in line with the SWC/DC model have been successful, and a partnership between LC and DC/SWC has begun. However, the transition requires structure, a communication plan, and leadership. Our project aims to build on the success on the LC movement[6], scale the workshops across the US, and formalize the partnership between LC community and DC/SWC organizations.

LC Module Development

The goal of LC is to provide librarians with high-quality skills training by librarians and for librarians. Our modules cover the full lifecycle of data-driven research. We do keep to the DC/SWC model by teaching hands-on training in data organization, management, and analysis to increase data literacy and improve research efficiency. However, the LC approach allows us to tailor the samples used during the lesson (datasets, content, and tools) to reflect the specific data and analysis needs of librarians, customizing knowledge frameworks familiar to librarians and motivated by real questions and samples relevant to the field of librarianship.

The original 2015 Library Carpentry event,[7] attended by 59 participants from 14 institutions,[8] taught materials initially developed by Dr James Baker, Owen Stephens and Daniel van Strien. Those modules were based on corresponding DC/SWC modules[9]. The original Library Carpentry modules were:

- Introduction, including jargon busting, data structures and regular expressions
- The Unix shell, including use of the command line and commands such as grep and sed to find data within files
- Git and version control
- Using OpenRefine for data clean up

However, these modules were drafts and not fully vetted. During a hackathon organized in June 2016, an engaged international community came together for a 24-hour sprint to transition the original teaching materials to Github, based on established DC/SWC practices. This resulted in consistent formats, improved lessons and a structure that was consistent with the SWC/DC model (See samples of LC modules in “Curriculum.pdf” attached to this proposal).

---

[6] Wilson, G. (2014). Software Carpentry: lessons learned. F1000Research, 3, 62. http://dx.doi.org/10.12688/f1000research.3-62.v2

[7] https://www.liberquarterly.eu/article/10.18352/lq.10176/

[8] The original workshop won the *2016 British Library Labs Teaching and Learning Award:* http://labs.bl.uk/British+Library+Labs+Awards

[9] http://librarycarpentry.github.io/outline/

Since then, LC modules have continued to expand with additional write-a-thon sprints targeted at improving the lessons. LC partnered with the Mozilla Global Sprint[10] in June 2017 to amend, update, and extend the existing Library Carpentry lessons and get draft lessons on on the following topics:

- SQL
- Python
- Web Scraping

As we look to serve the wider needs of librarian training, discussions have begun regarding adding future modules on software and data savvy approaches to librarianship. These would include new lessons on data librarianship and RDM, metadata skills, etc.

LC Instructor Training

In the past, formal DC/SWC instructor training was a barrier to entry for many librarians because the cost of one training session was too high and not geared to the librarian community. Without instructors in the LC community, the approaches between LC and DC/SWC remained segregated. Therefore, CDL sponsored a LC focused instructor training course held in Portland, OR in May 2017. This was the first time the DC/SWC community is training new DC/SWC instructors in direct support of the LC movement. This event was a huge success, drawing 98 applicants for 28 available slots.[11]

This also offered a new vision for how LC could transition from a grassroots group to one fully integrated into the DC/SWC processes for guaranteeing quality at future workshops. We intend to offer additional instructor training courses in all regions of the US to ensure this partnership will prosper.

LC Workshops

Library Carpentry was initially an informal, one-time event held in London. However, since then, there have been dozens of additional workshops held in Australia, South Africa, United States and more. Since July 2016, LC workshops in the US, averaging

[10] Mozilla Global Sprint: https://mozilla.github.io/global-sprint/projects/ and project page for the Library Carpentry project: https://github.com/data-lessons/librarycarpentry

[11] https://datapub.cdlib.org/2016/12/08/announcing-instructor-training-for-librarians/

[5] L. Barone et al. (2017), "Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators," *bioRxiv* 108555, doi: https://doi.org/10.1101/108555

[6] National Academies of Sciences, Engineering, and Medicine (2016), *Future directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020* (Washington, DC: The National Academies Press), doi:https://doi.org/10.17226/21886

40 attendees, have occurred in San Diego,[12] Berkeley,[13] Pittsburgh,[14] and Boston.[15] Additional workshops are planned now that we have formally certified LC instructors.

3 Our Approach

Our project centers on the creation of a new position, Library Carpentry Project Coordinator, to facilitate necessary work across the LC community. With the leadership of this new position, our project will sponsor additional DC/SWC instructor training courses to produce additional cohorts of certified LC trainers, coordinate efforts to enrich existing LC modules, develop new lessons on significant topics in data and software literacy for librarians, and coordinate national planning for conducting LC workshops. We will also provide much needed technical and organizational resources to act as a catalyst for ensuring that the US LC movement is maintained and prospers.

The outcomes from this project are:

- Build an effective, low-barrier, scalable way for librarians to learn and apply new skills. Bridge the LC community with the pedagogical infrastructure shared with DC/SWC and provide more opportunities for crossover for mutual benefit between librarians and researchers
- Partnering with professional associations, we will hold 6 DC/SWC instructor training sessions across the USA and online, particularly in underserved communities and regions (Southeast, Northeast, Midwest/Great Plains). The resulting nationwide network of over 300 librarians certified as DC/SWC instructors that set a foundation for future LC networks
- Recruit LC Project Coordinator to coordinate the instructors, training sessions, manage the collaborative changes to lessons, and assess the impact of our work. With this support and the efforts of certified instructors, we will strive to scale to over 600 workshops each year (2 workshops per year per instructor)
- Continue the work of the LC community to complete and formalize of the set of 7 LC lessons customized from DC/SWC modules to train librarians in data and software skills
- Finalize the list of and complete a set of 3 new LC modules created to train librarians in software- and data-based based approaches to data librarianship and RDM, metadata skills, etc.
- Link our US effort with existing global effort such as participating in 2 annual Mozilla Global Sprints to create and improve LC lesson modules

[12] https://ucsdlib.github.io/2016-07-18-UCSD/
[13] http://www.tim-dennis.com/2016-10-10-UCB
[14] https://datascienceinlibraries.github.io/2017-05-18-pitt/
[15] https://weaverbel.github.io/2017-05-11-lc-boston/

- Build plan for long-term LC efforts including instructor training, module development, LC training sessions, and assessment
- Creation of a LC advisory board of 4-6 community specialists to help transition LC from its current state as a grassroots effort to a more formal, sustainable relationship with DC/SWC

4 Project Design

The project outcomes described above have been organized into five focus areas: governance, lesson development, instructor training, assessment, and outreach.

Governance and organizational structure

In a volunteer community where participation is based on passion and desire to contribute to the greater good, a solid organizational structure is crucial to achieve deliverables, sustain engagement and communication, and interoperate within a larger international community. We will transition LC from a grassroots initiative to a sustainable component of the US's national digital platform. Therefore, our project activities will be guided by the recruitment of a new Library Carpentry Project Coordinator and an external advisory group.

A new Library Carpentry Project Coordinator position will be created at CDL to support these efforts. A draft job description is included in the "Project Staff" section of this proposal. This new position will be a full-time employee focused on organizing nationwide efforts and would be recruited from existing LC and DC/SWC communities.

In addition to the new Project Coordinator position, an external advisory group will meet quarterly and be comprised of 4-6 prominent DC/SWC and LC exponents, as well as library leaders. The goal of this group will be to ensure the project executes on this grant proposal, to guide the actions of the LC Project Coordinator, and to build a plan to transition to a more formal, sustainable relationship with DC/SWC by July 2019.

To this aim, we have received initial commitments from both Tracy Teal (Executive Director of Data Carpentry) and Jonah Duckles (Executive Director of Software Carpentry) to participate on the advisory group.

- **Tracy Teal** is co-founder and the Executive Director of Data Carpentry. She received her PhD in Computation and Neural Systems from California Institute of Technology and was an NSF Postdoctoral Researcher in Biological Informatics. She worked at Michigan State University as a Research Specialist with the Institute for Cyber-Enabled Research and then as an Assistant Professor in Microbiology. She is involved in the open source software and

reproducible research communities, and is an editor for the *Journal for Open Source Software*. Online CV: https://www.linkedin.com/in/tracy-teal-059136b/

- **Jonah Duckles** is Executive Director of the Software Carpentry Foundation. He has had various career trajectories including long-term strategic planning, landscape ecology research, monitoring of global agricultural production and supporting the computational workflows of researchers. He now helps research organizations build capacity for data-driven analysis by spreading the practices of Software Carpentry around the world. Online CV: http://www.jduck.net/cv/

In the beginning months of the project, we will formalize additional advisory group members from the library and training communities.

Training module development

The success of this project, and the LC movement, will be determined by the quality and quantity of its curriculum. As part of the efforts of the new LC Project Coordinator, we will create an online catalog of training modules and formalize a process for open nomination and prioritization of new module topical areas. We will gather feedback from the existing LC community and from the newly constructed advisory board. Once priority is established, we will optimize existing modules, create new lesson modules for data and software skills training, and create new ones for library-specific skills (For samples of LC modules, please see “Curriculum.pdf” attached to this proposal).

All LC training modules will be published under permissive CC-BY licenses and given DOIs for persistent citation; librarian authors will be able to add modules to their academic portfolios for evaluation during tenure or promotion review. The LC Project Coordinator will stay apprised of developments in tools and technology and ensure that needed new lessons are developed and taught as part of the evolving LC curriculum.

To help facilitate module development, the LC Project Coordinator we will oversee ongoing collaborative writing sessions as well as 2 “write-a-thons” during the aforementioned Mozilla Global Sprints. Work during these sessions will include:

- working on new incubator lessons
- incorporating an existing lesson into the Library Carpentry suite
- bringing together lesson threads developed by different community members
- developing a workflow for monitoring upcoming workshops and publishing our list of events in a public place
- adapting our lesson materials based on feedback from assessment surveys
- developing instructor guides for lessons

With this process, we will work to complete and formalize of the set of 7 LC lessons customized from DC/SWC modules to train librarians in data and software skills. These

could include existing DC/SWC modules on jargon busting, working with command line tools, basic data manipulation, scripting, data quality assurance, and statistical analysis. In addition, we will also finalize the list and complete 3 new LC modules on training librarians in software- and data-based based approaches to librarianship. These could include topics such as data librarianship and RDM, metadata skills, etc.

Instructor training

The two most significant questions Library Carpentry is currently facing are: “Who are the instructors?” and “How do I become one”? Establishing a pipeline of librarians who can teach both researchers and fellow librarians technical skills, and encourage them to go on to become instructors themselves, will be a key component of sustaining the US LC community. We will maintain and publicize a directory of certified instructors, and distribute a periodic newsletter with instructor and event opportunities.

We will encourage growth of the US LC community by quickly expanding the roster of certified trainers. To begin with, we will follow up with the 28 librarians who went through recent CDL-sponsored instructor training in May 2017,[16] and support them in their endeavor to teach their first workshops. An assessment of their experience will be useful in shaping subsequent training opportunities, as will consideration the methods used by DC/SWC.

The curriculum borrowed from DC/SWC for train-the-trainers events has solid roots in research on best practices in education and aims to equip the future instructors with practical skills to effectively teach students who may come from different backgrounds and have different research goals. Train-the-trainers programs are run either online for six weeks, incorporating self-study work and regular bi-weekly conference calls, or during intense two-day face-to-face events.

LC will focus on holding 3 face-to-face events and 3 six-week online course during the grant cycle. Based on current attendance expectations, we expect to grow our instructor community to 300 by the end of the grant period. We will focus on underrepresented areas such as Great Plains/Midwest, Southeast, etc. in making location and timing selections. Training will be open to librarians representing the full range of academic libraries: research, Baccalaureate, and community college.

Workshops and Assessment

The above efforts by the LC Project Coordinator and wider LC community expand opportunities for scaling LC workshops for librarians. As new LC instructors are certified, new networks and regions will be staffed to host workshops. The LC Project

[16] https://datapub.cdlib.org/2016/12/08/announcing-instructor-training-for-librarians/

Coordinator will organize and monitor these efforts. With our goal of certifying 300 new instructors during our grant period, we will need to scale to handle 600 LC workshops across the US per year. These workshops are customized for the audience and, since we are certifying all instructors, can include a mix modules from LC or DC/SWC. These workshops typically happen over one- to two-days with continuous education opportunities available via online chat rooms and in-person communities.

With 600 workshops a year and and average attendee list of 40, we can expect to reach 24,000 attendees to workshops conducted by our instructors. This type of scale is impossible for one person to organize alone. This is where the networked structure of DC/SWC is helpful. Their processes for self-organizing workshops will offer a scalable model for the LC community. As we scale, our effectiveness will be monitored through transparent assessment processes. We will leverage prior work done by DC/SWC.[17] Following the practice established by DC/SWC, we will rely upon pre- and post-workshop surveys for immediate assessments of satisfaction, as well as collecting data on the long-term impact of the workshops.[18] At the beginning of our grant period, we will compile previous LC assessment metrics to build a benchmark for future ongoing LC assessment work.

Outreach, training coordination, and communication

We will build communication channels and connections to all levels of the nationwide library community. Libraries are as diverse as the populations they serve, but there are areas of commonality and providing training and instruction is one of those areas. While continuing to use open, transparent platforms (e.g., Git, Gitter, Etherpad), the LC Project Coordinator will foster key relationships with institutions and people willing to commit to sustained workshops and instructor training for librarians. In order for this to happen, Our project will develop well-documented, responsive channels of communication (via etherpad, gitter chat, Library Carpentry website, github community repository, and email newsletter), a flexible and scalable workflow for lesson development, and a path for instructor certification that parallels that already established by DC/SWC. As our goal is to complement and extend DC/SWC efforts, we will maintain contact and coordination with those communities.

We will also partner with major library and research professional societies and initiatives (e.g., ALA, SLA, ACRL, RDA, Force11, RDAP, IASSIST, CJCS, Diversity Alliance) and offer co-sponsored LC workshops. In addition, we will identify key institutions dedicated to the LC mission and willing to support a core set of LC instructors. Ideally, these institutions will already have affiliated DC/SWC instructors

[17] http://www.datacarpentry.org/assessment/
[18] https://doi.org/10.5281/zenodo.165858

willing to collaborate with librarians certified in LC. We will strategically locate workshops and instructor training to reach populations and communities not currently served by readily available skills training. Our goal is to build a truly national network for training librarians that continues to thrive beyond this grant period.

5 Diversity plan

LC is valuable to all librarians. However, it is most applicable to those librarians that are working with research libraries and those working with data/software. We will enlist the ACRL Community and Junior College Libraries Section (CJCS) and Diversity Alliance to help ensure participation in LC training opportunities by librarians from all sectors, particularly underserved and underrepresented institutions and populations. Our long-term goal is for the LC network to reach all corners of the US. Toward this aim, we will also focus our 2-year project on building a foundational network of instructors in diverse regions of the country. While we will rely on expertise at research institutions, we will hold training events across multiple geographic regions and ensure that LC remains responsive to the needs of underserved libraries.

6 National impact

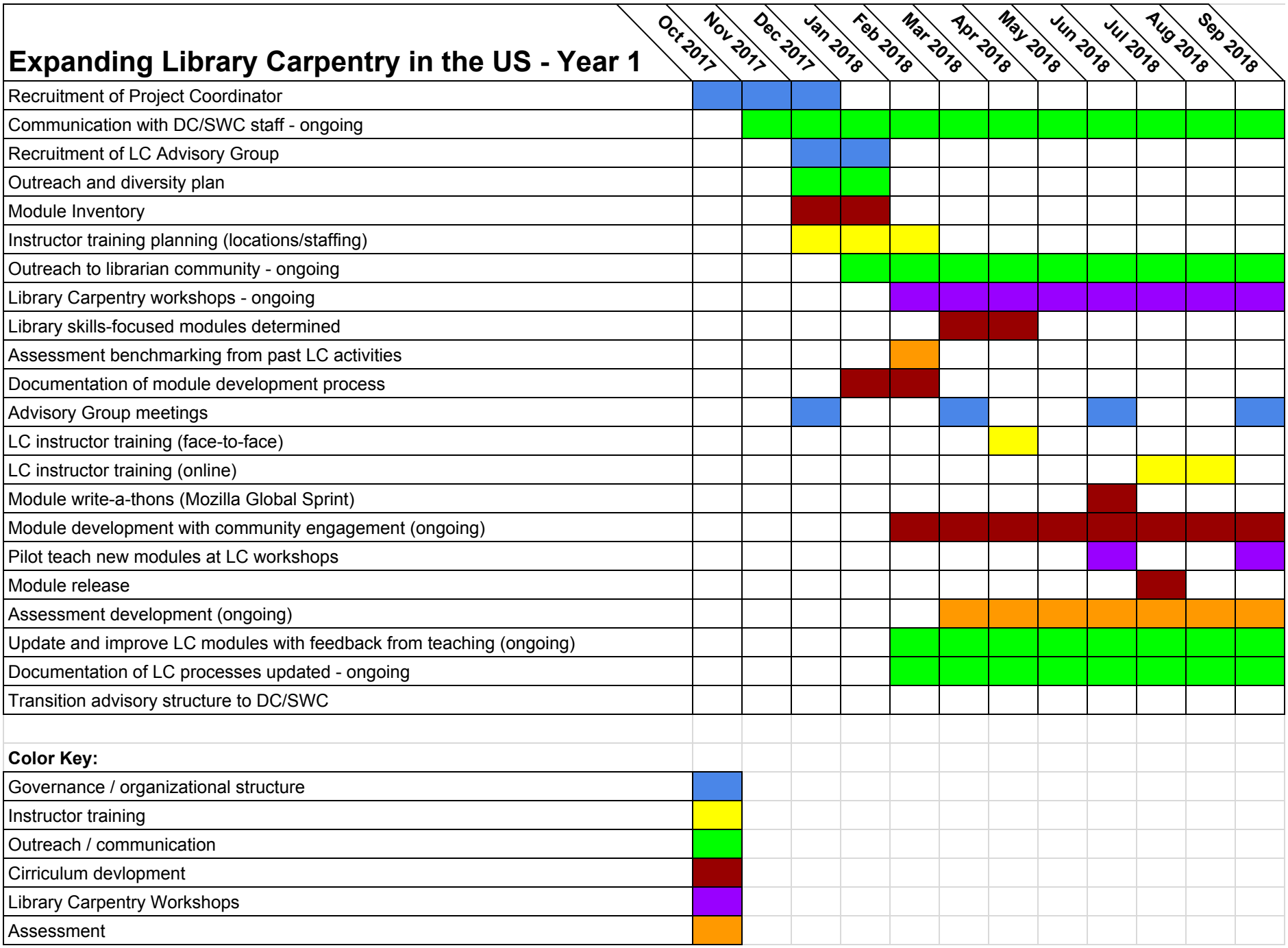
Surveys of DC training participants revealed extremely high levels of satisfaction and impact,[19] which are anticipated for our efforts:

- Workshop participants expressed more confidence in working with data. 99% continued to be more confident after 6 months.
- Participants are exploiting newly-learned skills more frequently 6 months after the workshops, and those skills, and the underlying tools, are helping them to improve their personal productivity.
- More than 50% of the participants have adopted improved data management and project organization, and many are reusing and sharing code or data.
- Many participants continue their learning with self-guided study or by participating in additional training courses.

Widespread US adoption of LC training will provide information practitioners with the skills needed to facilitate and enhance exemplary national stewardship of library collections, as well as ensuring the continuing relevance of libraries as strong anchors for supporting research and scholarship. Our efforts will support the IMLS through curriculum development, inclusive training opportunities, and support for communities of data and software literacy practice.

[19] https://zenodo.org/record/165858

Expanding Library Carpentry in the US - Year 1



Expanding Library Carpentry in the US - Year 2

	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019	May 2019	June 2019	Jul 2019	Aug 2019	Sept 2019
Recruitment of Project Coordinator												
Communication with DC/SWC staff - ongoing												
Recruitment of LC Advisory Group												
Outreach and diversity plan												
Module Inventory												
Instructor training planning (locations/staffing)												
Outreach to librarian community - ongoing												
Library Carpentry workshops - ongoing												
Library skills-focused modules determined												
Assessment benchmarking from past LC activities												
Documentation of module development process												
Advisory Group meetings												
LC instructor training (face-to-face)												
LC instructor training (online)												
Module write-a-thons (Mozilla Global Sprint)												
Module development with community engagement (ongoing)												
Pilot teach new modules at LC workshops												
Module release												
Assessment development (ongoing)												
Update and improve LC modules with feedback from teaching (ongoing)												
Documentation of LC processes updated - ongoing												
Transition advisory structure to DC/SWC												

Color Key:

Governance / organizational structure
Instructor training
Outreach / communication
Cirriculum devlopment
Library Carpentry Workshops
Assessment

DIGITAL PRODUCT FORM

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

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

PART I: Intellectual Property Rights and Permissions

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

For copyrightable works created in the performance of the scope of work, ownership will either vest with the creator or The Regents, however, we plan to license all copyrightable works created in the performance of the scope of work via a creative commons license, requiring only attribution

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.

For copyrightable works created in the performance of the scope of work, ownership will either vest with the creator or The Regents, however, we plan to license all copyrightable works created in the performance of the scope of work via a creative commons license, requiring only attribution

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.

NA

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

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

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

We will create lesson cirriculum (7-10) as well as web-
based help files and resources and social media communication (Twitter, blog)

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.

All required equipment is currently owned by CDL and will not need to be purchased.

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).

We will build cirriculum with github using markdown. We will offer web-based training information and project updates using standard web conventions (HTML, CSS, etc.).

B. Workflow and Asset Maintenance/Preservation

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

Our lessons will be tested using best practices of the Software Carpentry pedagogy. Once approved by our advisory group, we will continue to monitor assessment feedback given after each lesson. At subsequent writing sprints, we will revisit and iterate further.

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).

We will manage all lesson development through github. When modules are finalized, we will publish outputs to CDL's digital preservation repository, Merritt. This will issue DOIs for quick, consistent access and citation.

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).

We will produce DataCite metadata for each published lesson. This will include Title, Author, Abstract, Version Number, and relationship types.

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

We will be using github to track the progress of our project. With each update and pull request, we will be able to track the evolving lesson development. We will then extract snapshots of the metadata at the time each module is vetted and launched.

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).

Each lesson will be issued a DOI and metadata will be submitted to DataCite. Widespread discovery will be available via DataCite metadata search index, which is fed into mainstream and library-focused data search indexes (Google, Share, Clarivate Data Citation Index, etc.)

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).

All curriculum will be available in an open github repository (https://github.com/data-lessons). All published lessons will be archived in Merritt, CDL's open digital repository (http://www.cdlib.org/uc3/merritt/)

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.

Previous DC/SWC lesson have been made available via Zenodo. Example: https://zenodo.org/record/570048#.WTWigsbMxE4

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.

NA

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.

NA

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.

NA

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

NA

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

NA

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

NA

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

NA

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.

NA

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

NA

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.

We will create sample datasets to be used as test cases in lessons. These will be examples for instruc tional purposes only.

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?

No

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).

No

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.

NA

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).

All datasets created during our project will be associated with a lesson. We will generate these with t he aim to build the most appropriate test cases for each lesson plan.

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?

All datasets created during our project will be associated with a lesson. The cooresponding lesson m odule will offer documentation on the ways each dataset should be used.

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

We will deposit all lessons and cooresponding datasets into Merritt, CDL's digital preservation reposi tory, as they are being created and vetted. We will then publish all modules as they are formally laun ched.

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

Name of repository: Merritt

URL: http://www.cdlib.org/uc3/merritt/

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

Our project aims to build new lessons and iterate on them until we are successful in training librarians in new data and software skills. We will revisit this plan continuously, as we execute on new lesson d evelopment.