Empowering Librarians to Support Digital Scholarship Research: Professional Development Training on Text Analysis with the HathiTrust

Eleanor Dickson Koehl, Harriett Green, Amanda Henley, and Terese Heidenwolf

Introduction

Librarians increasingly support and engage in digital scholarly research on academic campuses. But this repositioning of librarians' work raises several challenges. First, "digital scholarship" encompasses a wide-ranging and evolving set of research behaviors, methods, and outputs across academic disciplines. Training librarians to become active facilitators of or partners in digital scholarship requires acknowledging the multifaceted ways in which digital research methodologies are applied. Next, amidst this shifting landscape, the profession has tended to separate digital scholarship or digital humanities librarianship into a professional sub-specialization. As a result, knowledge of key skills, tools, and strategies for digital scholarship are not widespread within academic libraries despite the growing demand for librarians to be trained in these areas.

“Digging Deeper, Reaching Further: Libraries Empowering Users to Mine the HathiTrust Digital Library Resources” (DDRF) was a project funded by the Institute for Museum and Library Services under the Laura Bush 21st Century Librarian program that aimed to lessen some of these challenges and provide digital scholarship training for librarians. The primary goal of the project was to develop and implement a training curriculum for librarians in the methods, tools, and concepts for text data mining, with the HathiTrust Research Center as the anchoring example of a text analysis research resource. The curriculum we designed for the project addresses a range of digital scholarship competencies and identifies points of convergence with expertise librarians already have. The three-year project (2015–2018) entailed two years of iterative curriculum development followed by a one-year workshop roadshow hosted by libraries primarily in the United States. We released the training materials as an Open Educational Resource in fall 2018. Throughout the course of the project, we gathered feedback data that we used to improve the curriculum and to assess our results.

This paper reports the outcomes of the DDRF project. We review the major milestones accomplished, including the collaborative design and development of a curriculum for training librarians in text analysis. We discuss the pedagogical approaches we employed to carry out a national series of training workshops, and the lessons we learned in digital pedagogy and professional development for digital scholarship. Finally, we share

* Eleanor Dickson Koehl, Digital Scholarship Librarian, HathiTrust, University of Michigan, efdkoehl@hathitrust.org. Harriett Green, Associate University Librarian, University Libraries, Washington University in St. Louis, harriett.green@wustl.edu. Amanda Henley, Head of Digital Research Services, University Libraries, University of North Carolina at Chapel Hill, ahenley@email.unc.edu. Terese Heidenwolf, Director, Research and Instructional Services, Lafayette College Library, Lafayette College, heidenwt@lafayette.edu.
findings from an assessment study that demonstrate librarians’ needs for practice-oriented, hands-on training about data-driven research.

**Literature Review**

A number of studies in recent years have addressed the development of digital scholarship service models in libraries. For example, reports from Ithaka S+R have looked at how libraries are accommodating their spaces, services, and personnel to provide digital scholarship support, and how to strategically support these needs. Among other such research, the 2016 Association for Research Libraries SPEC Kit #350 report on Digital Scholarship Centers in libraries found that support is relatively new: Of the 73 responding libraries, 67% of the digital scholarship library personnel had joined within the past five years and 74% had been doing digital scholarship work in the library for five years or fewer; and digital scholarship support is frequently distributed across the library from librarians and archivists to professional staff and student assistants. Two recent volumes from ACRL and Purdue University Press present a variety of use cases of digital humanities programs and projects in libraries. Groups such as the ACRL Digital Scholarship Section and the new Libraries Special Interest Group in the Association for Digital Humanities Organizations are among the growth areas for librarian communities involved in Digital Scholarship. These studies and initiatives show that support for digital scholarship in libraries is still nascent, but growing rapidly.

The 2017 Coalition for Networked Information (CNI) and EDUCAUSE report on “Building Capacity for Digital Humanities” notes that to establish robust programs for digital scholarship and data science, there is a need for librarians who, more than simply support but actively engage with researchers. Achieving that potential is a people problem: How can more librarians of all career-stages be empowered to engage in data-driven research in the diverse learning and research environments they encounter? Several libraries have launched programs to train their librarians in digital scholarship skills, such as Columbia University Libraries’ “The Developing Librarian,” and the Harvard Library Lab, among others. Additionally, over the last five years, a number of initiatives and projects have sought to close the gap between librarian expertise and desired skills in digital scholarship through workforce development and training programs. These programs include the Rochester Digital Humanities Institute for Mid-Career Librarians, the Data Visualization Institute at North Carolina State University, and the ARL Digital Scholarship Institute.

Other research has explored professional development needs for librarians. Projects such as Data Science in Libraries have recently released findings on how to train librarians in data science work. Educopia’s “Mapping the Landscapes” study also offers insight into the current state and expected needs for professional development in librarianship. Our study from the DDRF project reveals further information about librarians’ needs within the context of a specific training opportunity that uncovers rich detail about preferred learning styles and topics of interest.

**Project Design and Methods**

The DDRF project team included librarians from five partner institutions: the University of Illinois, Indiana University, the University of North Carolina at Chapel Hill, Northwestern University, and Lafayette College. These colleges and universities demonstrate institutional diversity in terms of their size, geography, and library-based programs for digital scholarship. The HathiTrust Research Center is co-located at the University of Illinois and Indiana University, and all project partner institutions are members of HathiTrust. While the project was rooted in HathiTrust, we developed the curriculum to cover a range of topics, and to be accessible to librarians regardless of their institution’s affiliation with HathiTrust.
The project consisted of three phases that roughly correspond to the years of the grant. In the first phase, we conceived of and drafted the first iteration of the workshop materials that we piloted in the libraries at Illinois and Indiana in spring 2016. Using comments and suggestions from the assessment form administered after each workshop, we revised the curricular materials following our iterative design methodology. In the second phase, we taught additional pilot workshops at all of the partner institutions and continued to make iterative modifications to the curriculum based on assessment data, feedback from instructors, and guidance from our advisory board. In the third phase, the project team taught a series of seventeen national and regional workshops using our near-finalized curriculum. The workshops were hosted by academic and research libraries around the United States, as well as one workshop in Kuala Lumpur, Malaysia in conjunction with the International Federation of Library Associations and Institutions (IFLA) conference. Finally, the project team packaged the curriculum into a freely-available Open Educational Resource that we released in fall 2018.

Alongside the workshop roadshow series, we began an IRB-approved research study in the third year of the grant. The study included an online survey that was used in place of the original post-workshop assessment form and a series of interviews, primarily with workshop attendees. The survey and interviews addressed both the workshop style and content as well as more general topics, including professional development preferences and needs in the LIS community and digital scholarship programs. The interviews lasted approximately 45 minutes, and they were recorded and transcribed. We analyzed both the interview transcripts and free-text responses from.
the survey in ATLAS.ti. Using a qualitative analysis method based in grounded theory techniques, the team developed an initial set of codes derived from themes they identified in the transcripts and responses, and then they coded each interview and set of survey responses multiple times to guarantee inter-coder reliability. Additionally, the project team enlisted the services of the University of Illinois’s Survey Research Lab to perform a quantitative analysis of the survey results.

Attendees of the workshop series self-identified with a variety of roles within the library profession, and while they represented a range of tenures in librarianship, they tended toward mid-career professionals. As reported in the survey, the most highly represented librarian role for attendees was Subject Specialist or Liaison Librarian (36.1%), followed by Digital Scholarship or Digital Humanities Librarian (14.4%), Reference Librarian (9.9%) and Technical Services or Metadata Librarian (9.9%).

The vast majority of respondents fell outside of the new professional category, with 36% reporting they had been in the library field three to ten years, and 28.9% reporting they had been in the field ten to twenty years. The workshops were free to attend. Registration was limited to librarians and was left open until full. Approximately twenty-five learners attended each of the day-long workshops.

### TABLE 1
**Librarian Role**

<table>
<thead>
<tr>
<th>Role/Position (n = 252)</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Specialist or Liaison Librarian</td>
<td>36.1%</td>
<td>91</td>
</tr>
<tr>
<td>Digital Scholarship or Digital Humanities Librarian</td>
<td>14.3%</td>
<td>36</td>
</tr>
<tr>
<td>Reference Librarian</td>
<td>9.9%</td>
<td>25</td>
</tr>
<tr>
<td>Technical Services or Metadata Librarian</td>
<td>9.9%</td>
<td>25</td>
</tr>
<tr>
<td>Library Administrator</td>
<td>6.3%</td>
<td>16</td>
</tr>
<tr>
<td>Special Collections Librarian or Archivist</td>
<td>6.0%</td>
<td>15</td>
</tr>
<tr>
<td>LIS Faculty or Educator</td>
<td>2.8%</td>
<td>7</td>
</tr>
<tr>
<td>Copyright or Scholarly Communication Librarian</td>
<td>2.0%</td>
<td>5</td>
</tr>
<tr>
<td>Data Services or Data Analytic Librarian</td>
<td>2.0%</td>
<td>5</td>
</tr>
<tr>
<td>Instruction or Information Literacy Librarian</td>
<td>2.0%</td>
<td>5</td>
</tr>
<tr>
<td>Library Assistant/Technician</td>
<td>1.6%</td>
<td>4</td>
</tr>
<tr>
<td>Library Developer</td>
<td>0.8%</td>
<td>2</td>
</tr>
<tr>
<td>LIS Student</td>
<td>0.8%</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>5.6%</td>
<td>14</td>
</tr>
</tbody>
</table>

Subject Specialist or Liaison Librarian (36.1%), followed by Digital Scholarship or Digital Humanities Librarian (14.4%), Reference Librarian (9.9%) and Technical Services or Metadata Librarian (9.9%).

FIGURE 2
**Number of Years in Library Field (n=253)**

<table>
<thead>
<tr>
<th>Years</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 3 years</td>
<td>15.4%</td>
</tr>
<tr>
<td>3-10 years</td>
<td>36.0%</td>
</tr>
<tr>
<td>10-20 years</td>
<td>28.9%</td>
</tr>
<tr>
<td>More than 20 years</td>
<td>19.8%</td>
</tr>
</tbody>
</table>
Pedagogical Approaches and Lessons Learned

Curriculum Design

The project began with an instructional design process that defined the learning goals and objectives for the curriculum based on the expected audience and desired outcomes. Relying on our prior experience supporting digital scholarship and examples found in library literature, we identified a set of key skills and areas of knowledge for librarians to engage in text analysis research. While we made some refinements to these learning objectives based on feedback and teaching experiences, our original set of core competencies formed a baseline that we referenced throughout the iterative development process to ensure the curricular materials were consistent with our goals.

We identified learning goals and objectives to address librarian-specific competencies that we felt were realistic to address in a one-day workshop. They centered on the unique role of librarians as resources, facilitators, and collaborators in the research process, and they prioritized expertise and experience already core to librarian-ship. As such, they fostered awareness of, and the ability to communicate about, text analysis research methods and techniques. Where the learning goals related to topics that were technical in nature, for example hands-on experience with the Python programming language, we emphasized exposure over mastery.

We aligned the learning goals to five training modules that follow the general workflow of a text analysis research project: an introduction, gathering text data, preparing text data, analyzing text data, and visualizing text data. For each module, we identified skills-based competencies that would be developed through hands-on activities. The activities we designed follow a sample reference question that threads through the workshop curriculum to provide a logical flow from module-to-module. Many, but not all, of the activities draw on tools and services of the HathiTrust Research Center. As previously mentioned, the activities deployed in several of the modules involve the participants executing Python programs, and it was important to our approach that we limit technological barriers encountered in such activities. As properly setting up a programming environment can take considerable time, especially in a workshop setting, we opted to utilize the web-based programming platform PythonAnywhere. We chose the modular format for the curriculum in order to divide each workshop into manageable, discrete units, and also to make it easier for others to pick-up and reuse the materials.

Workshop Outcomes

Overall, survey respondents rated the DDRF workshop highly, and 85.1% found the workshop either extremely or moderately useful. On a scale of 1 (not effective) to 100 (effective), the mean ratings for the effectiveness of each the workshop materials, activities, teaching methods and style, and organization were over 80.

Additionally, respondents reported a general shift toward familiarity with concepts such as the command line interface, Python programming, and web-based text analysis tools, with each of these concepts showing a downward shift in perceived discomfort from before the workshop to after the workshop, as well as an upward shift in perceived comfort from before to after. It seems there is an interest in additional similar trainings, as 96.8% of respondents would like to learn more about the topics presented in the workshops.

As it relates to the format of the workshops, we identified several key areas where attendees expressed the most and least satisfaction. For example, we found that pacing and organization were a concern for attendees. Likely as the result of anonymity, respondents tended to provide more critical feedback in the survey than in the interviews. In the survey, they expressed the most dissatisfaction when they felt the workshop had been rushed, too much content presented, or the set-up for activities disorganized. For example, when asked in the survey what they would like to add to or change about the workshop, one respondent wrote, “Workshop needed more time to properly cover all materials, and to allow attendees time to get setup for hands-on activities.” In fact, the
most highly requested improvements to the curriculum related to the ratio between the amount of content presented and the length of the workshop. We tried to improve on these pitfalls through the course of our pilots and the roadshow series, and there was a trend toward more positive feedback as the workshop series progressed.

Attendees were most satisfied with the content of the workshop and the way learners were provided the full curriculum, including the slides, handouts, and instructor script, for their independent study or reuse. One interviewee said, “I really, really appreciate and respect the way that these materials have been prepared and are designed to be reused, because if we wanted to do something like that here for our faculty, if we wanted to wrap something up here, we could.” They also tended to offer positive comments about the way the workshop presented real-world case studies and practical hands-on examples. Speaking about the theoretical concepts presented in the workshop, one attendee said, “Getting that explained and also applied to different use-cases was very helpful because sometimes it’s a little tricky when things are in the abstract to understand their applications, but the presenters really had some good example use-cases to talk about and that we could go see online too.” As we describe in more detail below, these comments align with what we heard from interviewees about their preferences for training styles.

**Digital Scholarship Training Needs**

In the post-workshop interviews and surveys, respondents expressed a clearly felt need for more training in technical skills both for themselves and for LIS professionals generally. The skills most frequently mentioned were programming or coding skills, particularly knowledge of Python; text analysis and text mining; and various data skills (data mining, data clean-up, data visualization). Interviewees also discussed the tension between needing to learn digital scholarship tools that may be soon replaced or obsolete. For example, one person said, “It is a challenge to know which [tools] to invest in, which will have lasting value, which won’t be around in a couple of years.” Still, interviewees generally felt that it was critical for librarians to invest time in learning new tools and methods, with one saying, “If we want to remain relevant to researchers going forward, those advantages certainly outweigh, in many cases, the time that it would take to work to learn” new tools.
A majority of respondents, however, did not think that librarians needed to develop in-depth expertise in highly technical topics, but rather need to be familiar with tools and trends to be able to make connections between researchers and experts. As one interviewee put it, “I don't think I necessarily need to be an expert in Python or other programming languages, but to build my skills enough to be able to at least troubleshoot or identify issues and at least—if I can't answer them myself—know other people who can answer those types of questions.” The survey asked respondents to choose all that applied from a provided list of ways they intended to make use of what they had learned in the workshop. 71.9% of survey respondents selected “as a starting place for learning more,” 58.6% selected to “teach/assist” researchers on their campus, and 38.7% selected to “partner with researchers.” These responses demonstrate the way that workshop attendees ranged from those who could envision themselves having an active role in digital scholarship to those who felt it was important to have a general understanding of the topic, trending toward the latter. We see a parallel here with the traditional idea that reference librarians need not be subject experts but need to know enough about reference sources to provide useful suggestions and referrals. One interviewee described this aptly: “When you work at the reference desk, you don’t necessarily need to be able to answer everyone’s questions but you can get them started. I think eventually it’s sort of similar in DH.” Rather than suggesting sources, librarians are now also needed to suggest tools and methods and make referrals to potential collaborators.

Although many respondents mentioned colleagues as resources for learning, including fellow librarians, others on their campus, and cohorts formed from attending a workshop or institute, or that they employed a variety of online professional development tools, many of them also saw real value to in-person workshops. They paint a fairly uniform picture of what they look for in such a workshop: an immersive experience—whether it be day-long or ten days—where the content is well structured, learning materials are available both before and after the workshop, and information is presented in “chunks” or modules interspersed with time for hands-on activities. Ideal hands-on activities allow participants to apply what they’re learning to real world problems or to learn from other students while instructors circulate around the room offering one-on-one assistance.

Conclusion

Through our iterative design methodology, which involved cycles of designing, piloting, assessing, and revising the DDRF curriculum, we were able to develop learning modules for librarians in text analysis that address current needs within the profession for digital scholarship training. Our assessment study uncovered the desire for other such opportunities. It revealed that professionals from across library organizations are interested in developing skills for digital scholarship. While some librarians, particularly digital humanities or digital scholarship librarians, feel they need advanced skills in technical topics in order to be effective in their work, most librarians who we interviewed felt they needed more generalized training in order to provide digital scholarship reference. Future training opportunities should focus energy on the workshop styles librarians prefer: those that are hands-on and practically-oriented. They should offer content most relevant to librarians’ professional development needs, such as training devoted to working with data, and they should be suitable for all librarians working in a variety of areas, not just those specializing in data services or data management. Current efforts to provide professional development training for librarians are a step in the right direction, and our research indicates that librarians themselves are eager for training opportunities like these.

Acknowledgements

Geoffrey Morse, Angela Courtney, Ruohua Han, Rachel Blomer, and John Henry Adams contributed to research and data analysis for this paper. The DDRF project team also included Leanne Nay, Lijuan Xu, Ben Jahre, and Sarah Morris.
Endnotes


Bibliography

Green, Harriett, Amanda Henley, Geoffrey Morse, Angela Courtney, Terese Heidenwolf, Eleanor Dickson Koehl, Leanne Nay, Lijuan Xu, Ruohua Han, and Sarah Morris. “Digging deeper, reaching further: libraries empowering users to mine the HathiTrust Digital Library resources’ curriculum.” http://hdl.handle.net/2142/102049.