Closing the Digital Curation Gap: A Grounded Framework for Providing Guidance and Education in Digital Curation

Helen R. Tibbo and Christopher A. Lee; School of Information and Library Science, University of North Carolina at Chapel Hill; Chapel Hill, NC, USA

Abstract

Over the past decade, a significant gap has emerged between the dramatic progress of research and development on digital curation and professional practices of archivists, librarians, and museum curators. There are many viable applications, models, strategies, and standards for long-term care of digital objects. However, many institutions are either not aware of the options or do not currently have the ability to evaluate and implement them.

The Closing the Digital Curation Gap (CDCG) project aims to provide guidance for professionals in small- to medium-sized repositories. We have elicited requirements from users through focus groups and semi-structured interviews. We are developing Digital Curation Guides to assist information professionals who are new to digital curation activities. We report on findings from the focus groups and interviews that inform the design of the guides. We also place our findings within the context of other work including the Digital Curation Centre’s Digital Curation Lifecycle Model, the DigCCurr Matrix of Digital Curation Knowledge and Competencies, the Library of Congress’ Digital Preservation Outreach and Education (DPOE) activities, the Society of American Archivists’ (SAA) Digital Archives Specialist (DAS) Certificate program, the DigCCurr Professional Institute, the European Commission-funded DigCurV training program, and the emerging CDCG Guides.

Introduction

Over the past decade, a significant and troubling gap has emerged between the dramatic progress of research and development on digital curation, on the one hand, and professional practices of archivists, librarians, and museum curators, on the other. There are now many viable applications, models, strategies, and standards for long-term care of digital objects. However, many institutions with a mandate to do this work are either not aware of the options or do not currently have the ability to evaluate and implement them.

Many information practitioners, regardless of their job titles, are conducting digital curation activities in a wide range of repositories and institutions. Such activities include creation of high-quality digital surrogates and orinals; selection and acquisition of existing digital assets; creation of metadata to support discovery, management, interoperability and preservation; producing agreements between content creators and repositories; managing copyright and other intellectual property rights issues; file format identification and management; ensuring reliable storage; and migration of content over time. Often these are new tasks and processes for which current staff have little training or experience.

There is a need to identify specific tasks and develop clear and understandable guides to good practice for information professionals working in libraries, archives, museums and other information centers and repositories. This paper focuses on the Closing the Digital Curation Gap (CDCG) project [1], and related work based in the School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill (UNC-CH) and how these projects seek to fill the gap between the burgeoning area of digital curation research and development and the practices and approaches of information professionals faced with digital curation responsibilities.

Closing the Digital Curation Gap

The CDCG project seeks to fill this gap between research and practice by providing guidance resources for professionals in small- to medium-sized repositories. Funded by the Institute for Museum and Library Services (IMLS) [2] in the US and the Joint Information Systems Committee (JISC) [3] in the UK, we have employed user-centered design methods in which we have collected requirements directly from users through focus groups and semi-structured interviews. Our user-centered approach was achieved by grounding our tool development in the real experiences of people working in cultural heritage institutions. We are applying what we learned from these information collection methods to inform the design and implementation of online Digital Curation Guides to assist information professionals who are taking on new digital curation roles and responsibilities. This project is creating such guides along with other tools to support the cultural heritage repository community, and especially staff in small- to medium-sized institutions in the US and UK, through researched, realistic, practical, and accessible guidance and advice.

The CDCG collaboration – between SILS, IMLS, JISC, and the Digital Curation Centre (DCC) [4] in the UK – is serving as a locus of interaction between those doing leading edge digital curation research, development, teaching, and training in academic and practitioner communities; those with a professional interest in applying viable innovations within particular organizational contexts; and organizations charged with disseminating such innovation and best practices.

Research Approach

CDCG has adopted a variety of methods to identify the needs of working professionals. We have drawn from many previous studies that have elicited the state of practice and needs of cultural institutions in relation to managing their digital collections. Two studies that have provided a particularly strong foundation for our work are those of the Northeast Document Conservation Center (NEDCC) in 2006 [5] and Cornell University Library in 2005 [6].
**Focus Group Design**

The CDCG team (Heather Bowden, Cal Lee and Helen Tibbo) conducted a series of focus groups with professionals who are responsible for digital collections in libraries, archives and museums. We administered four focus groups that included a total of 25 participants. We held two focus groups in conjunction with the American Library Association (ALA) annual meeting: one group with six participants and one group with nine participants. We held one focus group in conjunction with the Society of American Archivists (SAA) annual meeting that was comprised of four participants. Our final focus group was held in conjunction with the Museum Computer Network (MCN) annual meeting and had six participants.

Participant recruitment was directed toward professionals who were responsible for the care of digital collections and had sufficient background knowledge on digital curation to contribute to the discussion. For the ALA and SAA focus groups, participants were recruited based on their previous participation in related professional workshops.

<table>
<thead>
<tr>
<th>Question Area</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Current Activity Currently Practiced</td>
<td>1A</td>
<td>2A</td>
<td>3A</td>
<td>4A</td>
</tr>
<tr>
<td>B. Tools &amp; Resources</td>
<td>1B</td>
<td>2B</td>
<td>3B</td>
<td>4B</td>
</tr>
<tr>
<td>C. Sources of Tools &amp; Resources</td>
<td>1C</td>
<td>2C</td>
<td>3C</td>
<td>4C</td>
</tr>
<tr>
<td>D. Needed Activities, Tools &amp; Resources</td>
<td>1D</td>
<td>2D</td>
<td>3D</td>
<td>4D</td>
</tr>
</tbody>
</table>

*Figure 1 - Question Matrix for focus group discussions*

Each focus group began by asking the participants, “What kinds of activities do you believe are involved with digital curation?” We then walked the participants through a set of high-level digital curation functions, asking the following questions for each function (see Figure 1):

- **What kinds of digital curation activities do you currently practice?**
- **For these activities, what tools and resources have you used?**
- **Where and how did you find these tools and resources?**
- **What other types of tools and resources would you find to be helpful?**

We then presented a draft mockup of a “decision tree” tool and elicited the group’s feedback on it, including how they might use it, and what they would see as the main opportunities or benefits and challenges of using it.

**Focus Group Results**

We identified many commonalities in concerns, priorities and resources consulted across the focus groups. These have informed our ongoing development of resources for professionals. While there are some very general resources that cut across librarians, archivists and museums (e.g. search engines and electronic mailing lists), participants expressed interest in resources that were particular to their institution types. This was particularly true for resources that serve as examples of professional practice. For example, a general summary of considerations related to a particular digital curation function may serve the needs of librarians, archivists and museum professionals, but a museum professional could be most interested in particular examples that highlight practices in museum settings.

Many of the focus group participants provided thoughts about terminological differences across the professions. Several observed that the language of digital curation tends to align quite closely with that of archival administration, so it might be easiest for archivists to approach, understand and apply. Participants in one of the focus groups remarked that the practices of libraries and archives are more similar to each other than either is to work within museums, suggesting that translating terminology across library and archives contexts is also likely to be less of an issue than translating into museum terminology. Two examples were potential problems that museum professionals might have with the digital curation literature’s use of the term metadata and referring to the materials in collections as “content.”

It was surprising how often the participants reported that their institutions were doing little formal digital curation planning. Many indicated that their institutions wanted to do more formal planning, but that most of their digital curation activities were performed ad hoc. One participant remarked, “Our plan is to come up with a better plan.”

Focus group participants revealed numerous challenges related to limited resources or support for their work. A theme that emerged in all of the focus groups was the difficulty of building and maintaining relationships with the IT staff in the participants’ institutions. The majority of participants were able to speak in detail about how they generate and manage metadata, but many expressed concerns about how expensive their metadata creation and management currently are. Preservation activities tended to be an area of great concern. Participants tended to be most uneasy during this part of the discussion and many were apologetic about current preservation plans and practices. Participants often conveyed the perception that there are not methods or tools available for them to adopt.

The focus group discussions about the decision trees were very useful to the CDCG project. Many of the participants indicated that these should not be static web pages, but should instead be built in a dynamic system that could be continually updated. They also indicated that they would like the ability to leave comments or even rate certain resources that were listed in the decision trees. As discussed earlier, several participants indicated that having information related to institutions of particular sizes or types (libraries, museums or archives) could be helpful.

**Getting Started Guides**

It became clear both from the focus groups and from our early efforts to develop the CDCG educational resources that true “decision trees” were not the most appropriate model to meet the goals of the project. A primary motivation for the CDCG project has been to offer assistance to professionals at the point in their work processes when they experience breakdowns or need for further insight. There is thus no obvious starting point for a decision tree that will be appropriate to everyone, and it is often
not a particular digital curation function that drives one’s search for professional guidance.

We have instead found that a useful organizing device is a common scenario that professionals are likely to face. Based on the CDCG focus groups, various forms of feedback from professionals who have taken part in the DigCCurr Professional Institute (their pre-institute survey responses, interactions in institute and what projects they have worked on), as well as many informal discussions with working professionals in many different institutions, we have identified five scenarios in which professionals could often benefit from guidance on how to get started. These are archiving websites, building institution repositories, caring for digitized collections, managing data (as opposed to text-based materials), and acquiring information off external storage media.

We have thus adapted the decision tree concept to instead develop a set of focused “getting started” guides. Because there was a consistent call for interactivity and adaptability in the CDCG resources, we decided to begin building it in the open source Drupal content management system and have it linked to the resources that we have already started collecting in the Drupal-based Digital Curation Exchange (DCE) website. Each guide is organized around seven main verbs, which we have borrowed from the work of the Library of Congress’s Digital Preservation Outreach and Education (DPOE) initiative: prepare, identify, select, store, protect, manage and provide. Under each verb, we provide one or more questions, e.g. “How should I prepare to archive web sites?” and “What do I need to identify in order to archive web sites?” These questions serve as hyperlinks and titles to more detailed pages that summarize main considerations and pointers to existing resources that the user might find helpful.

The CDCG Guides are works in progress. In order populate them with content that is both timely and attuned to specific practical considerations, we have drawn from a variety of sources. This has included a set of interviews with expert professionals.

Interviews
We have conducted a series of interviews with practitioners who have special expertise in the five areas identified above: archiving websites, building an institutional repository, digitizing collections, managing and preserving data sets, and acquiring digital information from physical media. In each case, we asked the interview participant:

- What is your job?
- How did you come to be responsible for [PARTICULAR ACTIVITY]?
- How did you first approach the work?
- [PARTICULAR ACTIVITY] involves a lot of different elements, and you can’t do them all at once. What have been your main priorities? How have you determined those priorities?
- What professional preparation have you drawn from in order to carry out [PARTICULAR ACTIVITY]?
- What resources have you relied upon?
- What are you suggestions for others who are confronted with [PARTICULAR ACTIVITY]?

All of the interviews have yielded valuable insights about how to approach the process, what questions to ask and what resources to consult. Two common themes have been the importance of planning and the value of professional conferences as venues for soliciting advice from colleagues.

Existing and Emerging Frameworks for Digital Curation Education
As a recognizable field of practice, digital curation is less than two decades old [7] with the origins of formalized instruction in this area dating to approximately 1997 [8]. With such a short history, no one educational framework or canonical set of courses, readings, or instructive exercises or internships yet exists for either graduate students or working professionals. Educators in the United States and around the globe are developing programs in response to current and projected needs in the information workplace. This paper focuses primarily on the curricular and training efforts aimed at working professionals.

Digital Curation Lifecycle Models
There are numerous lifecycle models to characterize digital curation activities and the journey of digital objects from inception to disposition. Some are very general repository models as in the OAIS Reference Model [9]; others attempt to visualize scholarly communications and the place of archived information [e.g. 10]. Some models seek to capture human workflows around data and its preservation [e.g., 11], and still others look at information and preservation from a user’s view [e.g.,12]. Among the most recognized of these visualization is the Digital Curation Centre’s DCC Curation Lifecycle Model [CLM] [13]. The DigCCurr project has been informed by the DCC’s CLM, along with the Australian archival Continuum Model [14].

Digital Curation Curriculum (DigCCurr)
The School of Information and Library Science at the University of North Carolina, Chapel Hill is engaged in digital curation education at the master’s, doctoral, and professional levels.

The DigCCurr (Digital Curation Curriculum) I and DigCCurr II projects have developed conceptual frameworks, educational offerings, professional field experiences, and research opportunities to prepare digital curation professionals [15]. DigCCurr I focused on developing a curriculum and practicum experiences for master’s students, while DigCCurr II is supporting doctoral and professional education, including summer institutes for continuing education of professionals [16].

As a cornerstone of the DigCCurr work, we have developed a 6-dimensional matrix for identifying and organizing the material to be covered in a digital curation curriculum [17]. A given curriculum unit can focus on a dimension in general or specifically as it intersects with one or more other dimensions. The Matrix is a tool for thinking about, planning for, identifying, and organizing the digital curation curriculum. It is also helping us to address the issue of core vs. specialized (optional) educational elements. We have found this matrix is equally useful for planning continuing education opportunities for information professionals as it is for developing graduate-level education.
Matrix of Digital Curation Knowledge and Competencies

The six dimensions of the DigCCurr Matrix are:

1. Mandates, Values and Principles. Core reasons why the digital curation functions and skills should be carried out and should serve as the basis for criteria to evaluate whether the digital curation activities have been carried out responsibly and appropriately.

2. Functions and Skills. "Know how," as opposed to the conceptual, attitudinal or declarative knowledge that dominates several of the other matrix dimensions.

3. Professional, Disciplinary, Institutional, Organizational, or Cultural Context. Understanding of challenges, opportunities and characteristics of particular disciplines or institutions (e.g., social science data archive in a university, commercial collection of scanned page images, state archives, serving a population with specific cultural norms)

4. Type of Resource. Types of resources that are the target of digital curation activities

5. Prerequisite Knowledge. Elements of knowledge that are instrumental to understanding and applying other aspects of the curriculum, including specialized terminology and characteristics of technologies

6. Transition Point in the Information Continuum. Points of transition that span from pre-creation design and planning all the way to secondary use environments.

As part of the DigCCurr Matrix (dimension 2), we have developed High Level Categories of Digital Curation Functions [18]. We have found the categories to be useful for planning and administering educational materials across the information continuum. Categories include areas such as: Systems Engineering and Development; Production; Selection, Appraisal, and Disposition; Identifying, Locating, and Harvesting; Transfer; Ingest; and Data Management [19]. We use this in development of our digital curation graduate courses, the SILS Graduate Certificate in Digital Curation, and the DigCCurr Professional Institute.

The DigCCurr Professional Institute: Curation Practices for the Digital Object Lifecycle, is designed to foster skills, knowledge and community-building among professionals responsible for the curation of digital materials. The institute was supported by IMLS Grant Award #RE-05-08-0060-08 at the School of Information and Library Science, UNC-Chapel Hill from 2009 – 2011. This year’s institute (2012) is being funded solely through participant registration fees. The Royal Library of Denmark has asked us to offer the institute for them in Copenhagen, which we will be administering in June 2012.

The Institute consists of one five-day session in May and a two-day follow-up session and a day-long symposium in early January of the next year. Each day of the summer session includes lectures, discussion and hands-on "lab" components designed to expose participants to a wide range of professional and strategic issues while providing sufficient detail on several topics to (1) leave participants with concrete ideas about next steps in their own work and (2) demonstrate the implications of the main ideas through specific characteristics of associated technologies. Much of the focus is on having participants situate themselves in the digital curation landscape and understand the interdependencies among workers across the digital curation lifecycle. We also include discussion and hands-on work exploring how to integrate various tools into a digital curation workflow. Participants leave the summer session with a work plan for the coming six months and return the following January to report on and analyze their accomplishments and obstacles to meeting their goals. The cohort approach provides an environment that facilitates participants keeping in touch and supporting each other in their work for years to come.

Digital Preservation Management (DPM) Workshops

Originally developed by Anne Kenney and Nancy McGovern at the Cornell University Library in 2003 and moved to the Interuniversity Consortium for Political and Social Research (ICPSR) of the University of Michigan, in 2008, the Digital Preservation Management Workshops “incorporate community standards and exemplars of good practice to provide practical guidance for developing effective digital preservation programs” [20]. This year, the Massachusetts Institute of Technology (MIT) Libraries will be the third host of this education opportunity aimed at managers engaged with digital curation. The National Endowment for the Humanities (NEH) has partially funded the workshops. Nancy McGovern has further developed this series of workshops since her move to ICPSR.

According to the DPWorkshop.org website, “the goals of the workshop are to foster critical thinking in a technological realm and provide the means for exercising practical and responsible stewardship of digital assets in an age of technological uncertainty.” Continuing this practical perspective, “the workshop sessions are geared towards making a digital preservation program doable for any organization and all of the sessions include as many relevant examples as we can fit” [20].

One of the hallmarks of this workshop is the Three-Legged Stool Model that illustrates that any successful digital preservation or curation program must have robust organizational and technological infrastructures, and possess requisite resources. Nancy McGovern has brought this model to the DigCCurr Professional Institute for which she serve as an instructor. As with DigCCurr, the DP Workshop participants develop action plans, in this case to-do lists for developing policies and workflows at home.

Digital Preservation Outreach and Education (DPOE) Program

The mission of the Digital Preservation Outreach and Education (DPOE) program of the Library of Congress “is to foster national outreach and education to encourage individuals and organizations to actively preserve their digital content, building on a collaborative network of instructors, contributors, and institutional partners” [21]. During the summer and fall of 2010, DPOE staff conducted a Training Needs Assessment Survey. The DPOE website states, “The most important result of the survey was the realization that there were very few professional development opportunities in digital preservation. Survey respondents identified a need for practical hands-on information and training to conduct and manage digital preservation” [22]. This baseline data provided DPOE with a starting point for a digital preservation education and outreach agenda.
Before conducting any training, DPOE developed the “Program Pyramid” model consisting of three levels (executive, managerial, and practical) with three types of educational materials and approaches (strategy support, project planning, and hands-on learning experiences) – one for each respective audience. In September 2011 DPOE conducted a nationwide train-the-trainer workshop at the Library of Congress.

**Digital Archives Specialist (DAS) Certificate Program**

In 2010, the Society of American Archivists (SAA) formed the Digital Archives Continuing Education (DACE) Task Force to develop “a detailed professional development curriculum on the subject of digital archives.” The DACE Task Force used the DigCCurr Matrix, along with the DPOE target audiences pyramid, to shape the resulting Digital Archives Specialist (DAS) curriculum [23].

The curriculum and the associated Digital Archiving Specialist certificate were created in response to one of the issues identified in SAA’s strategic plan: “Rapidly changing information technologies challenge archival principles, practices, and communication protocols, demanding effective leadership from the archives community to access, capture, and preserve records in all formats.” DAS is structured in tiers of study so that individuals can select courses based on their specific knowledge, training, and needs.

There are four tiers of study within the DAS program:

- **Foundational Courses** focus on the essential skills that archivists need to manage digital archives. Practitioners and those working directly with digital materials are the target audience.
- **Tactical and Strategic Courses** “focus on the skills that archivists need to make significant changes in their organizations so that they can develop a digital archives and work seriously on managing electronic records.” These courses are primarily aimed at managers.
- **Tools and Services Courses** focus on specific tools and services that archivists need to use for their work with digital archives. They are practical in nature and focus on specific software products and other tools and primarily are directed toward practitioner archivists.
- **Transformational Courses** focus on the skills that archivists need to change their working lives dramatically and transform their institutions into full-fledged digital archives. The main target audience is administrators.

**DIGCurV**

DIGCurV or Digital Curator Vocational Education Europe is a project funded by the European Commission’s Leonardo da Vinci program to establish a curriculum framework for vocational training in digital curation [24]. According to the project website, “DIGCurV brings together a network of partners to address the availability of vocational training for digital curators in the library, archive, museum and cultural heritage sectors needed to develop new skills that are essential for the long-term management of digital collections” [24]. DIGCurV has developed an Evaluation Framework to help individuals and organizations design, provide or assessing digital curation curricula. Building from its own environmental scan, the DCC Digital Curation Lifecycle Model, the DigCCurr Matrix, the DPOE target audiences pyramid, the Research Information Network’s Researcher Development Framework [26], and the DAS Curriculum, the DIGCurV Evaluation Framework has 6 areas:

- Area 1: Knowledge and principles
- Area 2: Skills and competences
- Area 3: Audience/profile types
- Area 4: Part of digital curation lifecycle
- Area 5: Teaching methods/training delivery
- Area 6: Professional context

DIGCurV is currently developing graphic models focusing on the skills, knowledge, and competencies required by the various digital curation audiences.

**Observations and Implications**

Reassuringly, there are numerous commonalities across the various initiatives to provide lifelong education for those involved with digital curation tasks and responsibilities. At the same time there are also many differences and some unique perspectives. Perhaps the most striking similarity across these projects and programs is the notion of audience. There is fundamental agreement that educational programs must be geared not only to experience, knowledge, and skill levels, but also to the specific roles and responsibilities that individuals assume. For digital curation there is no effective training program that takes a one-size-fits-all-approach. While this appears perfectly reasonable – specialized training for the breadth of professional positions – such specialization results in extensive complexity and the need for many training resources. Because digital curation involves work across the entire information continuum and around the DCC Digital Curation Lifecycle Model [13], training cannot simply be aimed at novice, intermediate, or advanced audiences as one might with teaching algebra or a foreign language. While there are overview workshops and webinars that attract many participants, beyond awareness building, courses focusing on specialized tools for specialized tasks are also necessary. Digital curation practitioners do everything from working with content donors, to developing metadata schemes, to disseminating content to designated communities. Because the tasks are diverse, so too must be the training to support these functions.

Each of the training programs also deal with a range of content focused on imparting knowledge and principles and developing skills and competencies. There is a clear recognition that those working in curation, especially practitioners and mid-level managers (often the same person in small organizations) not only need the background knowledge to be able to understand issues, develop policies, and implement effective and efficient workflows; they also need to have hands-on skills and understand what goes on “under the hood” in order to communicate and collaborate with technical experts and evaluate the products of their work.

Most of the programs also recognize the need to contextualize teaching within the participants’ own organizational contexts. Frequently training courses will ask students 1) to assess their institutions’ digital curation readiness at the onset of the workshop; 2) bring a problem to work on during the workshop; 3) prepare an action plan to work on once back in their institution; or 4) all of the above.

Archiving 2012 Final Program and Proceedings
One of the areas that all training courses and programs struggle with is the most appropriate delivery mechanisms and program duration. With tight budgets and overworked staff, not all institutions can afford to send employees to lengthy workshops and especially not several of these that may complement each other. Many potential students would like to see more webinars in which they can participate from their desks while also frequently asking for hands-on training. This raises questions about how much one can learn in a short period of time and through particular media of communication.

The UNC DigCCurr and CDCG projects have found that perhaps the most valuable combination of knowledge and skill is that of being a great communicator who understands enough about the processes and implications of digital technology and preservation workflows and standards while also understanding the language and mindset of content creators in order to work with this varied set of stakeholders. The role of digital curator (e.g., the person heading up a repository effort within a library) is an agent who must build bridges across the disciplines, curation functions, levels of the organizational hierarchy (line worker, manager, administrator) and around the curation lifecycle in order to ensure the long-term preservation of our cultural, scientific, and governmental heritage.

Conclusion

The CDCG Guides are designed as a starting point for individuals and organizations that have little experience with digital curation but are faced with developing curation practices and programs. While not a training program in and of themselves, we are building them to be a logical precursor to attendance at the training programs discussed above. They presently focus on five scenarios that individuals are likely to face in the workplace. We envision the guides developing over time and serving as both a starting point and reference tool to complement active training programs.

References

Author Biographies

Helen R. Tibbo, Alumni Distinguished Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill teaches in the areas of archival management, digital curation and preservation, appraisal, and reference. Her research explores best approaches in implementing digital curation practices and provision of digital curation education with particular focus on workflows of content creators. She is a Fellow and Past President of the Society of American Archivists.

Christopher (Cal) Lee is Associate Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches archival administration, records management, digital curation, and information technology for managing digital collections. His research focuses on long-term curation of digital collections and stewardship (by individuals and information professionals) of personal digital archives. He is particularly interested in the practical and ethical implications of digitizing tools and methods into professional practice.