The International Journal of Digital Curation
Issue 2, Volume 6 | 2011

Educating Digital Curators: Challenges and Opportunities

Peter Botticelli,
University of Arizona

Bruce Fulton,
University of Arizona

Richard Pearce-Moses,
Clayton State University

Christine Szuter,
Arizona State University

Pete Watters,
Arizona State Library, Archives, and Public Records

Abstract

This paper describes a number of critical challenges faced by digital curation educators and suggests how the choices we make in building educational programs may impact the development of curation as a professional discipline. We focus on curriculum and program building as key steps in defining the educational needs of curators, and we argue for greater collaboration among educators, researchers and practitioners in the field, as a way to speed the emergence of curation as a discipline and to foster the integration of curation programs within libraries and archives.[1]

[1] This paper is based on the paper given by the authors at the 6th International Digital Curation Conference, December 2010; received December 2010, published July 2011.
The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

In recent years, many universities in the U.S. have launched new courses and programs designed to prepare information professionals to curate digital collections. Often, such programs have been motivated by the perceived need for updated practices and skills within established professional fields, including Library and Information Services, Archives and Records Management, and Scholarly Publishing, among others. In general, the growth of educational programs for digital curators has been driven by a widespread assumption that institutions are now, or soon will be, shifting resources away from print-based activities, investing more in digital assets production and management. Funding agencies have also played an influential role in promoting curation education. In the U.S., the Institute of Museum and Library Services (IMLS) and the National Endowment for the Humanities (NEH) have explicitly tried to advance digital curation capacity-building as a priority for the information professions. Other organizations, such as the National Science Foundation and the National Institutes of Health, have followed similar priorities by funding cyberinfrastructure research and development, and by developing formal data management and data sharing policies for research projects.

With the rapid emergence of digital curation as a field of study, and the urgent need for education in this field, educators have found themselves addressing fundamental challenges as they seek to build programs and define the knowledge base and skill sets needed by practicing curators. Our aim in this paper is to outline some key strategic issues we face in curation education, drawing from our teaching and administrative experience gained by directing three different, yet complimentary educational programs with a strong interest in digital curation. One is the Digital Information Management (DigIn) graduate certificate program at the University of Arizona, an IMLS-supported program that began teaching in 2007. A second is the newly-created Master of Archival Studies program at Clayton State University (CSU) in Morrow, Georgia, and the third is the Scholarly Publishing (SP) graduate certificate program at Arizona State University.

The three programs discussed here serve different professional communities, and yet they share complimentary goals in preparing students for roles in digital curation. The DigIn program is designed to help a broad range of librarians, archivists and museum professionals adapt their existing knowledge and hands-on technology skills to the demands we face in curating digital collections. Peter Botticelli directs the program, and the other authors of this paper have been directly involved with DigIn as a student, a faculty member, and as advisory board members.

The Clayton State University program began its development in 2004, with CSU and the Georgia State Archives collaborating on a proposal for a master's degree in archival studies. In 2010, with support from a grant from the U.S. Department of Education, the CSU program was formally approved by the Board of Regents of the University System of Georgia and Richard Pearce-Moses was hired as its first director. The Georgia State Archives and the National Archives at Atlanta are adjacent to the CSU campus, offering unique opportunities for students to apply classroom knowledge by working with rich collections through hands-on projects and internships.

The Scholarly Publishing graduate certificate program is mainly designed to build careers in the publishing industry. The program has over a thirty-year history that focused on traditional print publishing studies, specifically editing and design. Christine Szuter became director of the SP program in 2009, following a decade as Director of the University of Arizona Press. Under her direction, the SP curriculum is being redesigned to meet the current needs of publishing in the digital world. Although students from any discipline can enroll in the SP program, many of its students are enrolled in the Arizona State University Public History (PH) program because the SP is also part of this program. The SP and PH programs both emphasize issues related to archival research, curation and digital skills.

Through numerous discussions in recent years we have found that, despite their differences, all three programs share many of the same basic challenges, particularly the need to adapt quickly to the profound, technology-driven changes many professional fields are experiencing today. We also see the emerging field of digital curation as a key source of insight into the demands facing professionals in libraries, archives and publishing houses. In responding to the current needs of students, as well as the demands of employers and funding agencies, educators face an immediate concern in developing rich, up-to-date courses for curation. Secondly, we face a related, yet broader challenge in building sustainable educational programs around digital curation. The remainder of the paper will address these two basic issues. We will also discuss the PeDALS digital preservation project, as a case example informing our work as curation educators.

Challenges in Teaching Curation

Defining the Curation Knowledge Base

In preparing course content for curation, we have encountered a number of practical challenges, the most obvious of which is simple editing; that is, deciding what content to *exclude* from our courses. In developing a course, especially in a new field, the weekly choices instructors make necessarily have the effect of setting intellectual boundaries or defining the subject from the students' point of view. As educators, it is clearly not enough for us to say that we will teach "everything we can fit in fifteen weeks." So, in deciding what to teach in a given course, we have to treat the subject matter at a level of abstraction sufficient to give it coherence and depth. Achieving this balance in a subject as expansive as digital curation—one that is routinely presented in very general terms *and* in granular detail for specialists—is difficult.

We can see this dichotomy quite clearly in the field's dominant heuristic to date: the information "lifecycle," which is at once an overarching, interdisciplinary view of information management, and also a set of specialized functions that must be carried out in their local context within each lifecycle phase. So, in a given course, is it necessary to teach the full lifecycle, or just selected parts of the lifecycle? If we choose the latter, is it necessary to offer a whole course on each phase, or would it be better to offer a series of "whole-lifecycle" courses that go into successively greater detail? This same heuristic applies to the field of publishing studies, which traditionally considers the lifecycle of the book, and now is more generally applied to the lifecycle of any type of published work in digital format. These are just a sample of the kinds of

questions educators face in developing curricula, and it's likely that different instructors and programs will reach varying conclusions.

In the DigIn program, for example, we decided to offer one course that essentially covers the "curate" half of the Digital Curation Centre's Curation Lifecycle Model[2], and a second course dedicated to the "preserve" half of the model. On the whole, the DCC model has proven to be a useful theoretical basis for choosing course content in DigIn, especially as it incorporates the essential information management concepts of appraisal and selection, which we see as a key point of continuity between the work of librarians, archivists and research data curators.

A second basic abstraction we have found useful in developing course content for DigIn is the concept of a "repository," which can be defined at a technical level as a system used to store digital content, but also as a set of policies or rules for managing content. Thus, by having students carry out hands-on exercises with repository applications (e.g. Drupal, DSpace) built on a particular system architecture such as LAMP (Linux/Apache/MySQL/PHP), we can present LAMP not as a system to be learned for its own sake, but as an abstraction – a model or metaphor for learning the variety of systems our students might need to use in the field.

A third heuristic we've used extensively in DigIn courses is the "information ecology" concept developed by scholars in social informatics and related fields, as a way to help explain the complex interactions we find in "systems of people, technologies, practices, and values." (O'Day & Nardi, 2003). In practical terms, analyzing technology and business practices from an "ecological" perspective means taking a holistic view, in which the institution's internal functions cannot be understood without reference to its external environment – surely a necessity in an age of networks and rapid innovation in technology? (Davenport & Prusak, 1997) One of the luxuries of teaching curation in a 15-week graduate level course is that we can devote the time needed to explain the theory behind system ecologies, something we could not hope to do in a shorter format. But given the time needed to work through this complex an abstraction, we gain a useful analytical framework that helps explain the design principles and negotiations involved in designing curation systems and services.

Similarly, in the SP curriculum the overarching concept of publishing is explored in depth. We ask what it means to publish or be published in a digital environment. Once published, how will that knowledge be accessed, discovered, and searched for online? Publishers, booksellers, libraries, wholesalers, and others organize print publications by "lists" or "fields," but what will those lists mean in a digital and transdisciplinary world, where the need for strong metadata will be required? By exploring these fundamental questions, our aim is to ensure that students understand how the curation of digital published works is a far more complex process than was the case in the print world.

In general, teaching digital curation means engaging with an immature discipline characterized by fluid professional boundaries and uncertainty in the development of vital infrastructure. The lack of established standards and best practices means that we cannot teach digital curation with the same level of specificity we have taught other

[2] DCC Curation Lifecyle Model: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

subjects. For example, traditional courses in cataloging focused on tactics – applying well-established rules and standards to collection items. By contrast, in teaching digital curation we must focus on strategic thinking and offer students a toolset they can use in varying ways to meet local objectives. In this context, our search for heuristics is not only practical – offering instructors a path through the fast-growing thicket of curation literature – but, ideally, it should also help inform our students' efforts to address the practical challenges they are likely to encounter as curators. For instance, we would like to think that an "ecological" approach to infrastructure design might help future curators negotiate the differences between, say, science data, digital humanities collections, and electronic records.

Besides the choice of abstractions and particular topics to include in courses, another key problem we face in selecting course content is the very practical issue of deciding which types or genres of literature to use. In an emerging discipline like digital curation, there is a natural tension between "research" publications (i.e., peer reviewed articles in leading journals) and "practitioner" literature that often comes in the form of project reports, conference proceedings and case studies. One might expect students to favor one category or another, but our experience suggests some dissatisfaction with both types. Students often complain about research articles being too theoretical, and about project reports being too narrow in scope. Both are seen as lacking context, which is not surprising given that curation literature is seldom written for a student audience.

At the same time, much of curation literature is grounded in particular research disciplines or in computer science, which can be challenging for interdisciplinary classes. In the DigIn program, instructors are often called upon to play a "translational" role in explaining how technical and discipline-specific concepts may be applied to the generalist problems we face in building information infrastructures. We also have to explain how long-established professional principles and values – representing the traditional functions of libraries, archives and museums – may or may not apply to digital curation. This is especially important given that roughly 60 percent of DigIn students work for U.S. libraries, in roles that range from traditional librarianship to archives. Our students report a growing concern with digital assets management across the library/archives community, with many institutions seeking to build or already having access to repository or content management systems. And yet, we've found that even among the 20 percent of our admitted students who work for larger academic and research libraries (specifically members of the Association of Research Libraries), relatively few have direct experience in curating digital assets across the lifecycle, as opposed to managing assets that represent digital surrogates for print collections owned by libraries themselves. Hence the translational aspect of DigIn courses, as we encourage students to examine their own experience – and especially the experience of their institutions – in light of the new functions associated with curation services.

Similarly, in the SP curriculum, instructors play a translational role in explaining the complex role of publishing in research universities, libraries and cultural institutions. Issues discussed include: the creation of research; the relationship of tenure, scholarship, and publication; the role of granting agencies in research; the changing role of libraries in the publishing field; and the impacts of commercial entities, such as Amazon, Google, and Apple on digital publishing and curation.

Another critical issue we face in selecting course content arises from the practical problem of keeping up with a rapidly developing field with an abundance of literature. The demands of pedagogy further complicate matters, as effective teaching requires a balance of depth, coherence, clarity and relevance of content to our students' career goals. In addition, because the field is in its incipient stage, many of the most current thoughts, ideas, directions, theoretical perspectives and research are found on a daily basis on blogs, websites, e-journals, listservs and other digital means of immediate communication.[3] Thus, the urgent need to remain up-to-date raises a dilemma for instructors. An analysis of readings in a DigIn course taught in fall 2010 (IRLS 671: Introduction to Digital Collections) illustrates the difficulty of teaching an immature discipline, even in a case where the instructor (in this case) actively tries to include a historical perspective. As Figure 1 shows, even with a conscious effort made to include "classic" literature, most of the assigned readings are recently published, and in fact, it would not be difficult to exclude all or most readings more than 1-2 years old.

pre-2000	2000-05	2006-10	Total
15	24	93	130
12%	17%	71%	100%

Figure 1. Publication dates for readings in IRLS 671.

However, in the rush to update readings it is possible that our curricula could reflect a bias toward short-term versus longer-term trends. Moreover, given the complexities we see in the curation process, we wish to avoid promoting a common fallacy historians of technology have much experience in debunking; namely, the idea that state-of-the-art practices and tools necessarily represent advances and not "reverse salients" in the overall effort to build infrastructures at scale (Hughes, 1983). For this reason, instructors have a great need for practical case examples highlighting the organizational and technical problems curators are seeking to address, as opposed to the particular tools or outputs of curation projects.

Learning Curation Skills

Apart from the knowledge base needed by curators today, educators face major challenges in helping our students acquire the practical skills they'll need for day-to-day work in curation. Not surprisingly, our students are strongly motivated by skills acquisition – more so, in many cases, than closing gaps in their knowledge of curation policies and techniques. Yet we've found that while most of our applicants have some idea of the skills they'd like to acquire, e.g. working with databases, applying XML markup to documents, they often lack a detailed understanding of how particular technologies are being applied by practitioners at the local level.

The lack of perspective on curation skills is not surprising given that, in the case of DigIn, we have found that our students tend to have little experience or prior training with the advanced tools needed in digital curation. In fact, while two-thirds of the students admitted to DigIn have graduate degrees, most are in Library and Information Services as opposed to more ICT-intensive fields. Even among the minority of our students who have experience in programming or systems administration, many have significant gaps in their skill set relative to the demands of curating data collections.

[3] See, for instance, the Digital Scholarship online publications authored by Charles W. Bailey, Jr. at: http://digital-scholarship.org/.

The same holds true for SP students, the majority of whom are recent humanities graduates. Although many of them can be described as digital natives, they have not fully embraced the world of digital information apart from their experience in research on humanities subjects. In many cases, they have not been directly involved in the creation and curation of digital scholarship. Of course, their limited training and experience in the digital world is understandable since the field of digital humanities is also in its infancy.

For DigIn, a further complication for skills acquisition arises from the expectation that many of our students will be changing jobs and seeking to specialize in different roles within the curation lifecycle. In this respect, we see curation curricula as somewhat analogous to many MBA programs, in which graduates expect to be prepared to take on a variety of jobs (e.g., finance, consulting or management), and also to change jobs regularly. However, DigIn students tend to be older and more experienced than a typical population of MBA students, as half of our students are over 40, and over half of our students have at least five years of professional work experience in libraries and archives. And yet, even among the 50 percent of our students who have five years experience or less, we find many students in need of substantial upgrading in their basic technology skills, as well as in their knowledge of curation systems and policies. Perhaps this is to be expected given the historically rapid rate of innovation in ICT generally, as well as in curation techniques in recent years. Still, DigIn students are very much at the forefront of a seismic shift in the information professions, as library and archival functions that have tended to evolve slowly over time are now facing the prospect of revolutionary change on a historically fast time scale.

In this context, in building a skills curriculum for curation, our goal for DigIn as well as SP has been to lay a foundation for continuous learning, helping students develop the capacity to specialize in particular curation functions, while at the same time maintaining a broad perspective on the information and publishing lifecycle. Above all, we expect our students to be able to communicate effectively across disciplinary lines. Thus, in dealing with a technological and organizational environment marked by disruptive changes, we have come to equate "skills" with the ability to solve complex problems by first identifying or defining a problem, applying potential solutions, and then evaluating and reporting the results.

Of course, curation involves a wide range of different skills, as indicated by the skills matrix prepared by the IMLS-funded DigCCurr project at the University of North Carolina. Figure 2 gives a simplified list of functions identified by DigCCurr.[4]

	Functions
1	Access
2	Administration
3	Advocacy and Outreach
4	Analysis and Characterization of Digital Objects/Packages
5	Analysis and Evaluation of Producer Information Environment
6	Archival Storage
7	Common Services
8	Collaboration, Coordination and Contracting with External Actors

[4] DigCCurr Functions: http://www.ils.unc.edu/digccurr/digccurr-functions.html.

9	Data Management
10	Description, Organization & Intellectual Control
11	Destruction and Removal
12	Identifying, Locating and Harvesting
13	Ingest
14	Management
15	Preservation Planning and Implementation
16	Production
17	Purchasing and Managing Licenses to Resources
18	Reference and User Support Services
19	Selection, Appraisal and Disposition
20	Systems Engineering and Development
21	Transfer
22	Transformation of Digital Objects/Packages
23	Use, Reuse and Adding Value to Accessed Information
24	Validation and Quality Control of Digital Objects/Packages
	Meta-level functions
25	Analysis and Documentation of Curation Functions
26	Education and Sharing of Expertise or Guidance on Curation Functions
27	Evaluation and Audit of Curation Functions
28	Research and Development to Support Curation Functions

Figure 2. DigCCurr functions and skills.

The DigIn curriculum was initially mapped out in 2006-07, while the DigCCurr matrix was being developed, at a time when little formal guidance was available for curricula in this area. Hence, in our initial run of the six DigIn courses starting in 2007, we focused on a set of broad themes or abstractions (as described above) as a guide for selecting course content. We also focused on developing hands-on exercises using common technologies in use by curators.

Having developed and continuously refined our curriculum since 2007, it's now apparent that our general approach to skills has been quite compatible with the analysis undertaken for DigCCurr. In mapping our current course content to the DigCCurr skills matrix, we find that DigIn covers 25 out of 28 total functions (27 if other SIRLS courses taught by DigIn instructors are included). Of course, in a program consisting of six 15-week courses, it's not realistic to expect our graduates to have a fully-developed skill set covering 25 or more curation functions, especially given our emphasis in DigIn on closely-guided instruction and hands-on practice. Hence our overall emphasis continues to be on building problem solving capacity and helping students acquire new skills on the job, as needed.

A key problem in teaching skills arises from the fact that in practical situations, skills are closely tied to individuals' performance of routines that are shaped heavily by the local context. As much as we try to simulate "real world" scenarios in devising course assignments, we expect students' actual work as curators to vary significantly from their coursework. These issues hardly negate the value of "hands-on" instruction, though. Actually, an educational setting might offer a better venue than a job for learning to deal with non-routine problems and in mastering unfamiliar technologies

and methods, in many instances. Simply put, as educators we encourage students to make errors and experiment freely with potential solutions to problems – activities that carry much higher risks and stresses for individuals on the job than in a classroom environment.

Another problem we have encountered in devising a skills curriculum for DigIn stems from the fact that curators are expected to solve a wide range of technical problems, as they work with varied systems and collections. Thus, in our analysis of the DigCCurr curation functions, we've classified the 28 functions according to three broad types of skills: "technical," "analytical," and "organizational." These categories are not mutually exclusive, as organizational skills may be needed to solve technical problems, and vice versa. Likewise, formal methods of analysis (i.e. of technical architectures or business processes) are often required to solve both technical and organizational problems. And yet, as we will argue below, each category represents a distinct set of problems and potential solutions, and so our concern as educators has been to ensure that our curricula offers a balance of skills appropriate to our students' needs.

To some extent, this "balanced" strategy might runs against our students' initial expectations, as many consider technical skills to be the most important and also the most difficult-to-acquire skills they need to curate collections. Hence students, at the beginning especially, often assume that the best strategy for learning curation would be to focus their attention narrowly on whatever tools are most often mentioned in job announcements. This approach is certainly important for continuing education workshops that are designed to meet immediate needs. But for programs with a longer time horizon, including ours, we can't assume that today's "hot" technology won't be replace in a few years time. Thus, if we choose to teach an XML schema such as MODS or PREMIS, for example, we need to ensure that the underlying skills gained by working with these schemas add value to a student's whole career, beyond helping them prepare for their first job interview after graduation. It is for this reason that we have focused on helping students acquire the full range of skills needed for curating collections and managing information systems and services.

Indeed, our analysis of the DigCCurr functions strongly suggests that students need a balance between technical and analytical skills. We refer to technical skills as "hard" skills that require individuals to perform discrete, specialized tasks using software and metadata. In each case, developing hard skills requires hands-on practice with syntax, whether in typing Unix commands or applying valid XML tags. Students who enter the program without advanced technical skills often lack the self-efficacy to approach unfamiliar technical environments with confidence.

Thus, in one DigIn course, students are given assignments that involve installing, configuring, and managing a basic LAMP (Linux/Apache/MySQL/PHP) computing environment. A key objective of these assignments is simply to give students confidence in working with a command-line interface and foundational application suite, thereby laying a foundation for advanced assignments with server-based repository systems like DSpace and EPrints, where maintenance and management tasks require a substantial degree of technical familiarity and also self-assurance in resolving problems.

In our hands-on assignments geared to learning hard skills, we explicitly assume that students' learning objectives reach beyond the particular tasks they're asked to perform, and beyond the actual systems. Below we give an example of a learning objective in our introductory Applied Technology course (IRLS 672):

> "Explore basic PHP syntax including common commands and functions, use of variables including strings, and simple looping and branching commands."

While the immediate learning objective for this exercise is specific to the PHP scripting language, the exercise is presented to students with an overarching context, in which we introduce a structured approach to the principles and practices of application development that carries over to any kind of computer language needed to work with a given system. Thus, in assigning exercises with PHP, our aim is not to teach PHP per se, but rather to help students understand the basic design principles underpinning the scripting components of a complex server environment and those related technologies that comprise a digital repository infrastructure. We find that by giving students hands-on practice with the basic commands and file structures used by the LAMP architecture, we can help students jump-start much additional learning with server based operating systems and applications essential for curation that are common across technology platforms and also across disciplinary domains.

We also find that the process of acquiring hard skills complements the effort to acquire analytical skills, which we treat as related to technical skills but operating at a higher level of abstraction, focusing on systems design issues. Thus, one might be able to analyze a database at a theoretical level without actually knowing Linux and MySQL commands, but in this case the two skill types are complementary. For this reason, in DigIn courses we match our more "technical" assignments (as in the PHP learning objective given above) with analytical tasks, and in working with systems, we make a point to include instruction with analytical tools used in technology project management and preservation repository certification, for example.

In general, analytical skills are a very prominent feature of the DigCCurr list of functions. By our reckoning, at least 15 out of the 28 functions require the use of some kind of formal analysis of the systems, business processes, metadata schemas and policies used to manage collections. Whereas the "Ingest" and "Destruction and Removal" functions might be primarily technical in nature, functions such as "Preservation Planning and Implementation" and "Selection, Appraisal and Disposition" clearly require thorough analysis leading to the development of formal rules and procedures. Thus, analytical tasks are a central feature of the DigIn curriculum, as we believe that repeated practice is as important to skill acquisition in this area as it is in developing the capacity for hands-on experience with with systems.

Finally, all of our courses include assignments geared to building organizational skills, which we define as the set of human-centered or "soft" skills needed to manage project teams and institutions (Pearce-Moses & Davis, 2008). Communication is the most essential soft skill, as curators and publishers have to be effective in negotiation, advocacy, contract management and leadership, especially as the information professions seek to foster a culture of innovation while maintaining a strong ethical grounding and institutional mission. In general, we view soft skills as being

fundamentally rooted in critical thinking, and in the attitudes and values needed to work effectively in the complex and politically sensitive area of curation policies. As with hard skills, we believe an academic environment has distinct advantages in helping curators to develop soft skills, particularly as we provide a low-risk environment in which students can share their analyses and ideas.

Program Development Challenges

Apart from the immediate problems we face in developing course content, we also face broader challenges in building sustainable programs of study around digital curation. As noted above, our ultimate goal is to ensure that our graduates receive credentials that will provide immediate as well as long-term value in the job market. To accomplish this objective, we have to appeal to a diverse body of stakeholders, including students, employers, university administrators, faculty, professional communities and funding agencies, as we seek to define those areas of competency that will be of greatest use to our graduates in the workplace. As presented here, the task of defining program competencies is closely related to the decisions we make in curriculum development, yet it takes a macro view of the overall learning outcomes envisioned for the program as a whole, as opposed to the micro-level objectives we set week-by-week for a particular course. Broadly speaking, a statement of competencies can be thought of as an overall "mission statement" for the program, setting out a combination of knowledge and skills our students should be able to demonstrate in practical settings after graduation.

For academic units, the task of defining competencies has major implications for a program's curriculum, as well as its administrative structure. In preparing the IMLS grant application for DigIn, for instance, SIRLS had to make a complicated decision on how it might (or might not) integrate digital curation with the School's M.A. program, which has a well-established curriculum that reflects our students' predominant goal of becoming librarians or archivists, as opposed to digital curators per se. SIRLS resolved this issue by deciding to establish a new program with its own credential: a Graduate Certificate in Digital Information Management, that operates in parallel with the M.A. Program, but with several courses available as M.A. electives. In choosing this route, the School signaled its intention to expand the boundaries of a traditional library/archives education, but not to redefine or replace the core elements of the existing master's degree, which is accredited by the American Library Association (ALA). Thus far, the decision to structure DigIn as a six-course Certificate and as a series of M.A. electives has been successful, with increasing numbers of M.A. students opting to enroll in DigIn as a secondary program, and with substantial numbers of applications by working professionals (the majority of whom are employed by libraries) interested in curation either as a career path in itself, or as a new set of competencies they see as enhancing their current professional roles.

In hindsight, after three years of instruction in DigIn, we have encountered remarkably little tension in our effort to market courses on digital curation to our current population of M.A. students, for whom the idea of working with digital data (as opposed to publications or print archives) is still a new and uncertain prospect. And yet, the interdisciplinary aspect of digital curation is well reflected in the set of competencies SIRLS defined for the M.A. program, as part of the ALA accreditation process completed in 2005. In fact, the M.A. competencies have proven to be a good fit with the learning objectives for DigIn. Thus, SIRLS calls for a foundational

competency in the "creation, organization, management, access, and use of knowledge and information," and with additional core competencies in research methods — matching our emphasis on analytical skills in DigIn — and also a basic competency in information technology that clearly matches our emphasis on hard skills in DigIn. In addition to the core competencies, SIRLS identifies a number of secondary areas of competency for M.A. graduates, including metadata, management and information ethics. SIRLS also emphasizes the need for students to gain "cultural competence," as the ability to analyze information needs and to serve users representing diverse cultural perspectives. At first glance, cultural competence might seem peripheral to digital curation, yet, in our framework, "cultures" are rooted in practical activities, as in the communities of practice we see directly engaged in curation. (Montiel Overall, 2009).

In general, the effort to define competencies has been an important factor in shaping education programs for archivists as well as librarians. The Society of American Archivists (SAA) first established guidelines for archival education in 1977. In 2002, SAA published "Guidelines for a Graduate Program in Archival Studies," and this document was revised in 2011 (SSA, 2011). The guidelines expect archives curricula to provide students with a knowledge base that includes all basic archival functions and their professional contexts, including the interdisciplinary aspects of archival work. The guidelines also provide a list of specific components that archives degree programs should include, covering both core and supplementary topics of importance to archivists.

Likewise, the Academy of Certified Archivists' Role Delineation Statement is a comprehensive and specific list of knowledge and skills that archivists should know. The statement organizes its competencies into eight areas, including: general knowledge; selection, appraisal, and acquisition; arrangement and description; reference services and access; preservation and protection; outreach, advocacy, and promotion; managing archival collections; and professional, ethical, and legal responsibilities (Academy of Certified Archivists, 2009). The full list of competencies ranges from theoretical foundations ("Archivists know and can apply the impact of social, cultural, economic, political, and technological factors on the evolution and characteristic of records and papers and their management") to the very practical ("Establish, maintain, and keep a record of communication(s) with creators and/or potential donors of records and papers").

As much as these lists of competencies aim to be comprehensive, none go into significant detail regarding the technical skills that archivists and librarians need to thrive in the digital era. Thus, in 2006, while president of the Society of American Archivists, Richard Pearce-Moses sought to better understand the skills that archivists needed to work with digital records. He organized a meeting of archivists with practical experience working with electronic records, in an effort to isolate particular skills they relied upon in their jobs. With the support of the National Archives and Records Administration, the Society of American Archivists, and the Arizona State Library, Archives and Public Records, some fifty individuals participated in the New Skills for a Digital Era Colloquium in 2006. Although led by archival organizations, the colloquium sought to understand the skills needed by librarians and records managers as well as archivists. A key finding by the Colloquium was that, as much as the profession needs a dramatic expansion of hard skills, archivists also need to improve their soft skills as organizations struggle to adapt to the demands of managing digital archives (Pearce-Moses & Davies, 2008b).

Today, most would agree that the shift to digital media has affected the basic tenants of archival theory, with the result that all major areas of archival knowledge, and the practical skills expected of archivists, need to be reconsidered. While the primary motivation for archival collections – preserving cultural memory, documenting individual and corporate activities, protecting rights, and ensuring accountability – remain essential in the digital era, the manner whereby these goals are achieved is likely to change substantially. Many time-honored techniques that worked with paper will simply not work with electrons. Paper may be transferred in boxes with little concern that the information will be altered in the process. Data may be transferred over the Internet, which introduces many opportunities for loss of integrity. Archival education programs need new criteria focused on the digital information ecosystem, thereby expanding the profession's established body of theory, knowledge and practical skills.

Indeed, the findings of the New Skills for a Digital Era colloquium, augmented by much additional discussion within the archives profession, became a primary influence in SIRLS's decision to establish the DigIn program and to seek a balance of hard and soft skills in the curriculum. Also, while DigIn has been cast explicitly as an interdisciplinary program, the archives perspective has remained an important influence in the course content (partly in response to a growing demand among SIRLS students for courses related to digital preservation and archives). Given the limited number of courses required for the certificate, DigIn can't serve as a substitute for a Master's degree in archives, as we cannot cover the full spectrum of competencies called for by the professional associations. Instead, DigIn has focused on helping students achieve competency in a few essential areas of strategic importance to the profession, especially digital preservation and digital assets management.

By contrast to DigIn, the newly-created Master of Archival Studies program at Clayton State University (CSU) has decided to emphasize digital archives, acting on the advice of the staff at the state and federal archives, in addition to numerous advisors. The program is based in the College of Information and Mathematical Sciences, rather than in the College of Arts and Science. The still-evolving curriculum for the program intends to provide coverage of all the competencies identified by the archives professional associations. It is describing additional, technical competencies based on the New Skills colloquium and the experience of practitioners (described below). Of course, even in a master's program involving 45 credit hours, it would be impossible to cover in depth all of the theory, knowledge and practical skills needed by archivists today. Thus, as with DigIn and SP, the underlying goal for CSU is to lay a solid foundation for continuous learning and the capacity to work effectively with digital archives.

At a minimum, students who graduate from the CSU program should be as comfortable with digital information resources as they are with traditional resources, such as paper and other analog media. The students should be familiar with the agents, activities and objects in the digital information ecosystem. As important, they should understand the interrelatedness of the systems used to create and manage digital records, especially as the manner in which records are kept in the office of origin may have a significant impact on the archivists' ability to appraise those records. Likewise, the format that the archivist uses to disseminate records may facilitate or hinder researchers' ability to work with the records. The technology the archivist uses to store

the records may make it particularly difficult or easy for future archivists to preserve the records, both in terms of the ability to render the records so that they can be used but also whether those future archivists can demonstrate the authenticity of the records.

In working with digital archives, CSU graduates should be able to perform many of the day-in, day-out tasks that archivists have done with paper records. They must be able to transfer digital records from the record creators to the repository. They must be able to survey and inventory records. They must be able to describe records, both as aggregates and at the item level, to facilitate access and use. They must be able to place records in secure storage and to maintain best practices for long-term preservation. In resolving more complex technical problems, they may need assistance from professional programmers. In these instances, graduates should have the technical knowledge to be able to explain the problem in appropriate technical terms, to evaluate the proposed solution to ensure that it is effective and sustainable, and to carry out tests to ensure that the resulting code works properly.

By exposing the students to "real world" technologies during their coursework, we expect them to gain a basic understanding of system components and how they interact. Minimally, they will be able to have an intelligent conversation with a technologist – an essential competency in itself. At the same time, we expect students to be able to accomplish basic curation tasks, including web markup, simple database design and queries, and setting policies for managing digital collections. Some of the more technically sophisticated students will have enough knowledge to continue their education and become proficient in programming, database design, or system administration.

On the whole, the CSU program seeks to give students a renewed understanding of archival theory in the digital era, which may, in fact, be the most important outcome of the program. DigIn has a parallel goal as an interdisciplinary program; namely, fostering collaboration across disciplines by showing points of commonality in the core principles we use to manage information. Indeed, we believe professionals in all fields need a richer understanding of how their professions and the materials they work with are being transformed by the emergence of the digital information ecosystem. Above all, our programs are designed to help students to get their bearings, survive and thrive in an era of rapid and profound change.

The Scholarly Publishing certificate program approaches digital curation from a different vantage point. With its focus on the creation of published works in print and online, the curriculum addresses how the world of research and scholarship have changed or are being effected by the digital world. How will publishers ensure perpetual access, whether paid for or open access, to digital publications? How will published works avoid a digital black hole where they cannot be retrieved because they were never properly curated? Students are trained in specific skills, such as editing, design, bookmaking, and related computer skills. They also learn how to develop business models, marketing programs and fundraising opportunities to foster sustainable organizations.

As in the case of DigIn, many SP students are currently enrolled in, or have, master's degrees in other graduate programs, primarily history, linguistics and English. Most are at the beginning of their careers. Their discipline-based knowledge, understanding of the construction of scholarship, and insight into higher education are critical in the field of publishing. Yet students are told that they are being trained for a world that does not yet exist. They must be prepared to actively define and shape the future of scholarly publishing, and to be responsible for establishing systems and effective organizational routines for curating digital publications.

In each of our programs, the essential technology competence we expect students to acquire is centered around problem solving – that is, coping with system breakdowns, troubleshooting, evaluating tools and systems in light of user needs, and finding workarounds as needed in the local context. In this regard, a key objective in building technical competency is to help our students overcome the "fear factor" that tends to arise when people are confronted by unfamiliar problems and systems.

The Value of Case Examples in Curation Education

Given the complex technical and organizational issues faced by curators in the field, we believe it is vital for curation educators to consult regularly with practitioners in developing case examples for use in classes, and in defining overall program competences. Since it was founded in 2007, the DigIn program has benefited greatly from the experience gained by members of our advisory board in carrying out research projects such as PeDALS (Persistent Digital Archives and Library System), and this knowledge base is now well represented in the CSU program.

In 2008, the Arizona State Library, Archives and Public Records received a grant from the National Digital Information and Infrastructure Preservation Program (NDIIPP) organized by the Library of Congress. With Arizona as the lead, other partners included state archives and libraries from Alabama, Florida, New Mexico, New York, South Carolina and Wisconsin. The project had four principal goals:

1. To articulate a set of business rules that make up the steps of curating a collection of digital materials, from acquisition, through description, storage, reference, access and preservation;

2. To test whether those business rules could be implemented in middleware to automate the curatorial process;

3. To test the use of LOCKSS as a storage solution;

4. To build a sustainable community of shared practice that would meet the needs of a wide range of repositories.[5]

As of 2010, the PeDALS project is nearing completion, with all the elements of the system having been developed sufficiently to demonstrate proof of concept. The partners have articulated general business rules for processing collections, including a shared schema for administrative, discovery and preservation metadata, and an administrative catalog to manage the process. At a practical level, PeDALS was able to implement business rules for automated processing of collections. At the outset, the

[5] PeDALS Project Objectives: http://www.digitalpreservation.gov/partners/states_az/states_az.html.

project assumed that the software implementation would need to be tailored for each record series. But over time, project staff were able to develop software that minimized, and sometimes eliminated, the need to customize the base code. Also, project staff found that middleware processing of collections was not as seamless as hoped, and yet, interruptions encountered in the workflow provided excellent opportunities for quality assurance.

The project staff also gained valuable experience by adapting LOCKSS to the demands of storing government records rather than serial publications, as it was originally intended to do. Making use of LOCKSS for archival collections meant solving a range of technical problems, such as the fact that number of potential government records could easily exceed the number of journal issues by several orders of magnitude, raising questions about the technical capacity of the underlying Unix file system. Furthermore, the potential size of the archives in bytes had the potential to overwhelm the LOCKSS polling mechanism. In fact, the project found a simple solution to the limits on the number of objects in the file system by encapsulating all the records from an acquisition into a single file.

Most importantly, PeDALS was able to demonstrate that it is possible to develop a community of shared practice. By working together, the partners were able to save the costs of developing redundant, shared systems. By using a similar methodology and technology, they were able to share code and reduce development costs. All the partners were hit very hard by the recession, which has raised some questions about the future of the project. The partners have a viable business plan, but the recession has strained even well-established programs to the breaking point.

From an educational standpoint, the experiences gained by the PeDALS team offer invaluable lessons for curation students, and yet it's often difficult to distill this kind of information from project reports and published articles that naturally tend to focus on the final outputs from a project rather than reporting in detail on lessons learned (shortcomings as well as successes) by project staff. The DigIn program has benefited greatly by working directly with advisors representing the Arizona State Library, Archives, and Public Records, making it much easier to incorporate the lessons of PeDALS and other digital projects in the curriculum.

Indeed, DigIn has benefited from the Library's work on two major projects, including PeDALS, which was preceded by research on ways to capture and preserve state agency publications from the Web. This work was carried out, in part, through the ECHO DEPository preservation project funded by the Library of Congress.[6]

In both projects, curators drew from a strong foundation in archival and library theory when setting objectives. Before the ECHO DEP project, librarians had been searching for individual documents and harvesting them manually for preservation – a process that clearly would not scale effectively. In addressing this problem, archivists were able to apply the principles of provenance and original order to reconceptualize the preservation work as archival, rather than bibliographic. In this view, websites would be selected according to their provenance – the agencies responsible for producing them – rather than by the contents of Web pages. By approaching state agencies' websites as archival collections, it was possible to appraise and select

[6] ECHO DEPository preservation project: http://www.ndiipp.illinois.edu/.

content as aggregates rather than items, and it was also possible to apply metadata to the documents based on the characteristics of the series to which they belonged (Pearce-Moses & Kaczmarek, 2005).

Of course, librarians and archivists working on the project needed a fair amount of technical knowledge to understand how technology might help them apply an archival approach to masses of Web-based documents. For example, to identify websites with relevant content, curators developed tools that parsed several thousand links on dozens of websites into a database, then queried the database for a distinct list of relevant Web domains. Given the scale of the collections involved, it was essential that the tools provided curators an efficient, systematic method with which to identify new content. Prototyping the tool to demonstrate the proof of concept required a knowledge of web architecture, Internet protocols, a simple scripting language, and SQL database queries. Note that it was necessary to employ a professional programmer to develop the tool.

Within the PeDALS project, curators had to develop rules to transform metadata from a variety of sources into a meaningful descriptive record. In some instances, curators had extensive experience writing MARC records based on Anglo-American Cataloging Rules (AACR2, 2005). Yet the rules provided little or no guidance for the problem at hand, as the records had no title page to transcribe. A curator commented: "You can't write rules to automate my job!" The curator failed to understand that the project was not trying to automate the job of writing bibliographic records and finding aids. It was trying to find new ways to describe materials. While AACR2 was of limited use, Charles Ami Cutter's 1876 classic "Rules for a Printed Dictionary Catalog" provided a practical theoretical basis for the project's metadata strategy. In fact, Cutter's basic objectives for a catalog were easily adapted to the digital environment, as they permitted someone to find a record when the provenance, title, or subject of a record is known (search); to show what the repository holds by author, on a given subject, or of a given type (browsing); and to assist in the choice of material as to edition and character. Similarly, his means suggested the minimum data elements necessary for a description: the author(s), a title, subjects, the form, and additional notes (Cutter, 1876).

In effect, the curator needed to find a way to map existing (and often limited) metadata to a title element, access points, and notes, enabling a patron to select records based on the description, rather than having to download each record. In the case of marriage certificates, these records included the bride's and groom's name, the date of marriage, and a license number. A meaningful title that could assist in selection could be formed by creating a string from the bride's name, the groom's name, the phrase "marriage certificate" and the date (e.g., "Jane Doe and John Roe: marriage certificate, 20 December 2007"). This, in essence, is an example of a business rule being applied across thousands of records in a single acquisition. Once a programmable rule for title creation is in place – a rule the curator can describe in plain language (e.g., "The item title should equal the bride's name + ' and ' + the groom's name + ':marriage certificate,' + the recording date") – the curator needn't worry about entering the title by hand for each item, as the PeDALS middleware could automate this function. Because the bride's and groom's names were provided in atomic form, access points could be formed by placing the last name first. Although the metadata did not include the location, all the certificates were from the Maricopa County Clerk, and the county could be listed as a location for all records by default. In general, theoretical

knowledge of cataloging effectively informed the process of mapping metadata to support discovery and access. The curator had to understand that their job was not to slavishly transcribe, but to provide information to help the patrons select materials relevant to their search. Practical knowledge of traditional methods of discovery and access were of limited value in this digital environment.

In carrying out the PeDALS project, curators required some very basic technical skills. They needed to know how use tools like tar or zip, and sftp or disk utilities to move files from the record keepers' offices. They needed to know how to check the integrity of the files using a hash value. In some cases they needed to know how to read a simple XML file to inspect the metadata, while in other cases they needed to be able to parse metadata in delimited flat files. PeDALS used LOCKSS for storing the records, and so curators needed to know how to install and configure a LOCKSS server, as well as how to write plugins to harvest the content.

As it happens, one curator who has worked on PeDALS (now as a Principal Investigator) was previously a DigIn student, and he was actually introduced to the project and wrote about it in a DigIn class, along with the Library of Congress NDIIPP project which funded PeDALS. The curriculum also covered a number of technologies and standards in use by the PeDALS team. These include the OAIS preservation standard, the TRAC certification process and the LOCKSS preservation system. DigIn also provided an introduction to metadata standards used to inform PeDALS' standard, including MARC, Dublin Core and PREMIS.

Also, the hands-on exercises in DigIn involving the LAMP architecture proved to be useful preparation for working with the PeDALS computing environment, which used different yet analogous technologies. Thus, PeDALS uses Microsoft SQL Server rather than MySQL, and a combination of ASP.NET and C# rather than PHP in the public dissemination Web interface. Drawing from his hands-on experience in DigIn, the curator was not intimidated by having to install CentOS Linux on a virtual machine to test the LOCKSS daemon and build specific LOCKSS plugins for PeDALS, which required skills not far removed from DigIn exercises with an Ubuntu Linux virtual machine.

Conclusion

In conclusion, the sustainability of educational programs for digital curation clearly depends on delivering credentials that will be widely recognized by employers, and so as educators we have an urgent need for greater clarity about the institutional and professional roles curators may be expected to play in the near and longer term. With the experience gained in projects like PeDALS, we are reaching a stage when educators can increasingly point to concrete problems and solutions faced by curators, with the result that we are shaping our programs around the maturing knowledge base and skills required for lasting careers with digital information. Of course, we continue to deal with serious challenges in developing and revising course content, and in building new programs to serve the whole community of information professionals. But, at the very least, we're off to a running start.

References

Academy of Certified Archivists. (2009) Role Delineation Statement. Retrieved May 10, 2011, from http://www.certifiedarchivists.org/get-certified/role-delineation-statement.html.

American Library Association, Canadian Library Association, and the Chartered Institute of Library and Information Professionals. (2005) *Anglo-American Cataloging Rules, Second Edition, Revision: 2005*. Retrieved May 10, 2011, from http://www.aacr2.org/.

Cutter, C.A. (1876) *Public libraries in the United States of America, their history, condition & management: Part Two. Rules for a printed dictionary catalog.* Washington D.C.: U.S. Bureau of Education.

Davenport, T.H. & Prusak, L. (1997) *Information ecology: Mastering the information and knowledge environment.* New York: Oxford University Press.

Hughes, T.P. (1983). *Networks of power: Electrification in western society, 1880-1930.* Baltimore: Johns Hopkins University Press.

Montiel Overall, P. (2009) Cultural competence: A conceptual framework for library and information science professionals. *Library Quarterly, 79*, 175-204.

O'Day, V. & Nardi, B. (2003) An ecological perspective on digital libraries. In A. Bishop, N.Van House & B. Buttenfeld (Eds.) *Digital Library Use: Social Practice in Design and Evaluation.* Cambridge: MIT Press.

Pearce-Moses, R. & Davis, S.E. (2008) *New skills for a digital era.* Chicago: Society of American Archivists.

Pearce-Moses, R. & Davis, S.E. (Eds.). (2008b) *New Skills for a Digital Era: Proceedings of a Colloquium Sponsored by the Society of American Archivists and the Arizona State Library, Archives, and Public Records.* Washington, D.C.: Society of American Archivists. Retrieved May 10, 2011, from http://www2.archivists.org/sites/all/files/NewSkillsForADigitalEra.pdf.

Pearce-Moses, R. & Kaczmarek, J. (2005). An Arizona model for preservation and access of Web documents. *DttP: Documents to the people, 33*(1), 17-24.

Society of American Archivists. (2011) Guidelines for a Graduate Program in Archival Studies. Retrieved May 10, 2011, from http://www2.archivists.org/gpas.