IJDC | *General Article*

Digital Curation Education and Training: From Digitization to Graduate Curricula to MOOCs

Helen R. Tibbo
School of Information and Library Science
University of North Carolina at Chapel Hill

Abstract

This paper traces the development of digital and data curation curricula. Due to the brief length of this paper, the focus is on North American initiatives and primarily on continuing education programs. It explores the strengths and weaknesses of professional workshops and the creation of graduate-level courses, certificates, degrees and MOOCs, as well as the role of funding agencies in this process. It concludes with an analysis of what is missing and what is needed to create the workforce required to steward digital assets in the foreseeable future.

Received 13 October 2014 | *Accepted* 10 February 2015

Correspondence should be addressed to Helen R. Tibbo. School of Information and Library Science, 201 Manning Hall, CB#3360, University of North Carolina at Chapel Hill, NC 27599-3360. Email: tibbo@ils.unc.edu

An earlier version of this paper was presented at the 10th International Digital Curation Conference.

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. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see http://creativecommons.org/licenses/by/2.0/uk/

Introduction

Curation of digital assets, whether cultural, educational, scientific or economic, is one of the central challenges of the early 21st century. (Beagrie, 2007; Bell, Hey, and Szalay, 2009; Digital Curation Center, 2005; DigitalPreservationEurope, 2007; National Science Board, 2005; National Science Foundation, 2006; National Science Foundation/Library of Congress, 2003; Waters and Garrett, 1996) The last decade has witnessed extensive progress toward robust repository architectures and cyberinfrastructure, (e.g., DuraSpace[1]; Fedora[2]; iRODS[3]; ISO, 2012b; National Science Foundation, 2003) preservation tools and strategies, (e.g., 4C[4]; Archivematica[5]; ArchivesSpace[6]; BitCurator[7];) and trustworthy and sustainable digital curation (ISO, 2012a). Key projects provide a foundation for ongoing research and development. (e.g., DataONE[8]; Planets[9] [now OPF[10]]). This work and the growing recognition that "one of the major challenges of this scientific generation [is] how to develop the new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams," led the National Science Foundation (NSF) in March 2007 to call for "developing a coherent data cyberinfrastructure in a complex global context" and a "national digital data framework" (National Science Foundation Cyberinfrastructure Council, 2007). This vision resulted in the NFS's "Sustainable Digital Data Preservation and Access Network Partners" (DataNet) funding program (National Science Foundation, 2007). Such a vision requires a cadre of digital curation professionals to work in libraries, archives, museums, data centers, and an increasing array of data-intensive organizations.

While the International Digital Curation Conference[11] (IDCC) is ten years old, the origins of digital curation education trace to almost a decade before that. In 1995 the Northeast Document Conservation Center[12] (NEDCC) conducted its first "School for Scanning: A Conference on Digitization, Microfilm, and Preservation," at the JFK Library in Boston, MA. (NEDCC, 2015). In 1996, Anne Kenny of Cornell University Libraries led "Digital Imaging for Libraries and Archives"[13] the first in a series of programs that would set the standard for early digitization workshops and lead to today's Digital Preservation Management (DPM) workshop series.[14] In 1998, the Humanities Advanced Information Institute[15] (HATII) at the University of Glasgow started a Summer School on Digitization for Cultural Heritage Professionals, under the direction of Seamus Ross. The HATII team taught this course in Europe many times and also at Rice University, 2000-2001, and with the School of Information and Library

1 DuraSpace: http://www.duraspace.org/
2 Fedora: http://www.fedora-commons.org/about
3 iRods: http://irods.org/
4 4C: http://4cproject.eu
5 Archivematica: https://www.archivematica.org/wiki/Main_Page
6 ArchivesSpace: http://www.archivesspace.org/
7 BitCurator: http://www.bitcurator.net/
8 DataONE: https://www.dataone.org/
9 Planets: http://www.planets-project.eu/
10 Open Preservation Foundation: http://www.openplanetsfoundation.org/
11 IDCC: http://www.dcc.ac.uk/events/idcc15
12 NEDCC: https://www.nedcc.org/
13 DILA: https://www.library.cornell.edu/preservation/dila.html
14 DPM: http://www.dpworkshop.org/
15 HATII: http://www.gla.ac.uk/subjects/informationstudies/

Science[16] (SILS) at the University of North Carolina (UNC-CH), 2002-2004.[17] While the focus here was on best scanning practice, the course was imbued with content on project management and technical standards. The lifecycle of content was discussed and users were factored into the scenarios. These programs were the start of what has developed into graduate and post-master's certificates, workshops of varying lengths and a wide range of foci, and most recently, MOOCs (Massively Open Online Courses) in data curation and management.

Nineteen years has witnessed a substantial growth in digital preservation and curation education, but existing programs remain insufficient for the training needs of the information and data management professions, let alone data creators, information technology staff, and other stakeholders in the digital curation lifecycle. Twelve years after the "It's About Time" report (NSF/Library of Congress, 2003), the need for education and training of digital curation professionals is more pressing than ever due to the massive increase in digital data. In 2010, OCLC published the report "Taking our pulse: The OCLC research survey of Association of Research Libraries (ARL) special collections and archives" (Dooley and Luce, 2010). This study surveyed the most pressing problems in special collections and archival repositories and found the "'most challenging issues' in managing special collections were space…, born-digital materials, and digitization." Clearly the need for graduate and continuing professional education in digital curation remains critical.

Several disciplines and professions have developed de facto practices and expertise in aspects of data management without guiding principles or an overarching vision of data preservation and reuse. Professional education for digital curation has generally involved on-the-job training and experimentation, possibly supplemented by workshops lasting a few days. (e.g., DCC[18]; DigCCurr[19], DigCurV[20]; DPC workshops[21]; DPM workshops[22], Library of Congress-DPOE[23]; NEDCC[24]; PTAB[25]; and UIUC[26]).

Professional workshops and graduate-level courses have grown up side-by-side, often taught by the same individuals and in part containing the same content; for current information professionals in one case and for future professionals in the other. The pedagogical challenges of these two educational strands, however, are quite different and extend well beyond just "what to teach."

Pedagogical Concerns of Professional Workshops and Continuing Education Programs

Professional workshop organizers must decide on the duration of individual events, delivery methods, pricing, and sequencing of multiple workshops into programs, as well

16 SILS: https://sils.unc.edu/
17 Digital Curation for Heritage Professionals – An Intensive Program: http://ils.unc.edu/images/
18 DCC – Digital Curation 101: http://www.dcc.ac.uk/training/dc-101
19 DigCCurr: http://www.ils.unc.edu/digccurr/
20 DigCurV: http://www.digcur-education.org/
21 DPC Training: http://www.dpconline.org/training
22 DPM: http://www.dpworkshop.org/
23 Library of Congress – Digital Preservation Outreach and Education: http://www.digitalpreservation.gov/education/
24 NEDCC – Digital Directions: https://www.nedcc.org/preservation-training/digital-directions/dd15/
25 PTAB – Courses: http://www.iso16363.org/courses/
26 University of Illinois Urbana-Champaign – Digital Humanities Data Curation Institute: http://www.dhcuration.org/institute/

as the content they will provide. Here chunking of content is significant as working professionals have limited time resources to devote to continuing or vocational education. Additionally, the question of whether such programs are sustainable looms large. This short paper affords only enough space to talk about some of the most important factors.

Audience

Before anything else, an institution or group setting out to provide continuing education must identify their target audience. Without this step there may well be no audience! Most of the content related to digital curation could be taught in some form to K-12 school children all the way to PhDs. It can be taught to those who need to manage their own content and it can be taught to those responsible for managing the digital content of others, such as data management librarians and repository curators. Needless to say, a one-size-fits-all message is not appropriate for all of these audiences. While everyone might well agree that one curriculum is not appropriate, it can be a daunting task to match what is taught to the needs of a specific audience. If the audience is heterogeneous, as is so often the case with workshop participants, curriculum creation become doubly difficult. The need to teach both fundamental principles as well as provide hand-on computer training can also be a confounding factor, as students often need to work at their own pace with the latter.

One of the best approaches to mitigate difficulties of mismatching content, depth of content, duration of instruction, and mode of instruction is to do a market survey before setting forth in workshop development. The Library of Congress, in conjunction with the NEDCC conducted just such a survey in 2010 before starting the DPOE program (Library of Congress, 2010) and the DigCurV project sent a large survey on digital curation education needs throughout Europe (Englehardt, Strathmann, & McCadden, 2012) as well as conducting a survey on available vocational educational programs (Karvelyte et al., 2012). Grant funded projects may conduct such surveys, both formally and informally, prior to writing the grant application that includes funding for the instructional program, thus no survey is part of their final grant products. Of course, many times the need to skill up a workforce that is on average decades beyond graduate education is obvious. What may not be so obvious is how to provide the instruction and at what cost. Market or profession-wide surveys can certainly provide data to assist with understanding audience needs and preferences.

What to Teach

The next pedagogical issue is, of course, what to teach. What should be the content of workshop? Will the focus be broad – a general introduction to digital curation – or will it explore more specified topics, such as curation of specific media? Will it look at digital curation as a whole or be limited to the management of scientific data? The answers to these questions must flow from the purpose of the workshop and audience needs. Most workshops arise from grant funding or institutions with a mission to provide continuing education to a specific profession or the public in general. If the institution is conducting the workshop on a cost-recovery or profit-making model then audience uptake is definitely a factor to consider in content to be covered.

Once the general topic(s) have been decided, the focus of each day or session must be selected. What, out of a range of items that can file an entire semester will be going into the workshop? Along with the selected content instructors need to consider the

 doi:10.2218/ijdc.v10i1.352

sequencing of the content and the placement of exercises, breaks, lunch, tours and demonstrations, and discussion, as well as lecture and slide presentations. What sort of social activities should be planned? Should all the "theory" and overview come on Day 1? Should lectures be interspersed with labs? Adult education generally views a mixture of activities as best, keeping peoples' minds fresh. Especially important is to provide active learning wherein students are asked to do something, maybe in a group, maybe at a computer, with short periods of lecture and slides. Just a warning – great exercises are much harder to make than great slides.

How Long to Teach

The next consideration is how long or short should the training be? Will this be a week-long event, three days, one day, a half day, or will it extend over a series of months? Any of these durations can produce great workshops. The important facets to juxtapose here are instructional goals, audience needs, and programmatic capabilities. If an institution can only host a day-long event then the planners need to consider that limitation and design a program that will provide worthwhile learning within that time frame. Cramming more into a time period than anyone can absorb is not a successful approach and is disrespectful of the audience. Often presenters feel that students can use extra slides on their own after the class. While this is true, an overly packed class often leaves students feeling overwhelmed and confused.

If an institution or project can afford to conduct longer training sessions then instructional goals and audience interest/capability to attend/pay come into play most clearly. Now the workshop goals must be in sync with "what the market will bear." Remember, there may be different or sub-audiences for generally the same content delivered over different time spans and through different modalities.

The week-long workshop has had a good deal of success in the digital curation arena (DigCCurr[27]; DPM[28]; UIUC[29]). A week provides enough time for students to become comfortable with each other, the instructors, and with the content. Students also have social time in which they can bond with their cohort and form a supportive social network. The DPM workshop has been running with this model since 2003.

Started as part of the DigCCurr II grant funded by the Institute for Museum and Library Services, the DigCCurr Professional Institute has taken a more extended approach to the week-long workshop. DigCCurr participants are sent materials in advance to read and are asked to post content about themselves on the DigitalCurationExchange or DCE. They come to Chapel Hill, NC for a week in May or June; stay in touch with each other and the instructors through various media and the DCE, and return again the next January to report to the cohort on the success of their projects that they conceptualized in the spring session. DigCCurr has been running since 2009 and has been self-sustaining since 2013.

What Educational Modality to Use/Where to Teach the Class

Today, face-to-face education is but one approach to teaching continuing education workshops. The benefits include possibly developing a cohort culture and fostering interaction among the participants and instructors. Additionally, some content,

27 DigCCurr: http://www.ils.unc.edu/digccurr/
28 DPM: http://www.dpworkshop.org/
29 University of Illinois Urbana-Champaign – Digital Humanities Data Curation Institute: http://www.dhcuration.org/institute/

especially that involving computer use, may be more easily done in a computer lab where the instructor can look over the shoulder of the student without a lot of special software and programming required. Face-to-face workshops may also be taken on the road to institutions (DigCCurr went to the Royal Library of Denmark and DPM has been taught in several institutions around the world) or be attached to conferences. Webinars are increasingly popular, as students do not have to travel to the workshop. Downsides may include participants working from their desks while listening to the content and a general lack of interaction.

MOOCs or Massively Open Online Courses are also appearing in the digital curation arena. University College London (UCL) offered a free Digital Curation MOOC[30] in May 2014 and UNC-Chapel Hill is producing one for Research Data Management for the CRADLE project in 2015[31]. At this point there is not enough data to judge the effectiveness of these efforts but MOOCs may prove to be highly effective in reaching large and geographically diverse audiences, such as researchers who need to learn to manage their data.

Sustainability

Perhaps the largest challenge for digital curation continuing education in the United States is that none of the curatorial professions (librarianship, archiving, or curatorship) require ongoing certification. Without a mandate to take classes and earn a certain number of continuing education credits every few years, workshop offerings often go half-filled and there is little control as to who is going to take a given course, as instructors are always looking to fill seats. It is hard for university programs, whose mandate is undergraduate and graduate education, to justify such workshops if they are not financially self-sustaining through registration fees.

The emergence of ISO #16363: Audit and Certification of Trustworthy Digital Repositories (ISO, 2012a) and the related standard ISO #16969: Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories (ISO, 2014), speak to the need for national bodies that will certify auditors and repositories. Movement in this direction may well increase the market for digital curation education and thus make it feasible for iSchools[32] to engage in more extensive professional education programs.

Some organizations, such as the Society of American Archivists (SAA) with its Digital Archives Specialist Certificate program,[33] are succeeding in making a profit from their continuing education programs. The Academy of Certified Archivists[34] (ACA) does require continuing education credits for its five-year re-certification program thus we can assume that some of the archivists taking DAS course are ACA members.

In order for digital curation continuing education programs to evolve beyond grant-funded project-based activities, professional organizations and academic institutions must develop business models and figure out how this activity will be sustained. Other questions include: how is the academic going to be rewarded for working in this arena beyond formal graduate and undergraduate education? How can we reach such a large audience (everyone needs to know about digital curation!)? And can iSchools provide

30 UCL: https://www.class-central.com/mooc/1960/introduction-to-digital-curation
31 CRADLE: http://cradle.web.unc.edu/
32 iSchools: http://ischools.org/
33 DAS: http://www2.archivists.org/prof-education/das
34 ACA: http://www.certifiedarchivists.org/

the digital curation teaching capacity that the government, military, corporate, scientific, academic and public sectors will require?

Professional Workshops – Conclusions

The issues discussed above are but a few facing workshop developers. Other challenges include: who should do the instruction? What training do instructors need? What is the best sequencing of brief workshops? Should courses have pre-requisites? How can the instructor best meet the needs of a diverse class? How much should courses cost? How can educators encourage institutions to support their employees in taking continuing education?

Graduate-Level Education

Limited, but quickly growing graduate educational opportunities in digital curation now exist. The University of Illinois at Urban-Champaign (UIUC), the University of Texas at Austin, the University of Michigan, and the University of Maine offer specializations or certificates. UNC-Chapel Hill has produced a master's level digital curation curriculum (Lee, 2009), a graduate certificate,[35] and a PhD-level curriculum. A review of a sample of top-ranked ALA-accredited information and library science (ILS) graduate programs provides evidence that while some schools offer courses in digital preservation and digital libraries, few make such coursework required, and even fewer offer much in terms of lifecycle digital object management content. Additionally, few faculty list "digital preservation" or "digital curation" as their specializations. UIUC, provides a concentration in data curation at the master's level (UIUC)[36] while UNC offers a 30-credit, primarily online post master's degree in data curation.[37] UNC-Chapel Hill is also in the midst of creating an entirely online Professional Science Master's in Digital Curation and is designing MOOCs on data management for researchers and information professionals (CRADLE).

Graduate programs are concerned not only with content but also with course sequencing, pre-requisites, and how courses fit with one another in an established university degree program. Perhaps the greatest challenge for graduate programs is to provide students with consistently meaningful internships and other hands-on experiences. Often students are limited geographically to working in repositories and organizations close to their schools; online students may have an even more difficult time in securing field experiences in which they will be mentored by an experienced and reflective practitioner.

Computer science programs as well as iSchools and Management Information Systems programs are all claiming territory in the data curation realm. While all these programs prepare people to manage, manipulate, and visualize data, only the iSchools thus far have taken an interest in the lifecycle management and preservation of the data. The iSchool programs have also taken the strongest interest in working with content producers and users across the content lifecycle.

35 SILS Certificate in Digital Curation: http://sils.unc.edu/programs/certificates/digital_curation

36 UIUC – Specialization in Data Curation: http://www.lis.illinois.edu/academics/degrees/specializations/data_curation

37 UNC – SILS Post Master's Certificate in Data Curation: http://sils.unc.edu/programs/graduate/post-masters-certificates/data-curation

It is expected that these digital and data curation programs will continue to grow in number and enrolment for the next few years, but ultimately, most of this content will just be folded into iSchool curricula. At some point in the not too distant future the digital domain will be considered the norm and distinctions such as "digital" curation will be dropped. It is important, however, that the notion of "curation" not be dropped. Never before in the history of librarianship or archivy has there been such a clear notion of the need to curate content over its entire lifespan. Certainly the highly volatile and fragile nature of digital files has pushed this notion forward, but it should not be lost as we become more adept in the digital domain. The foreshortened nature of preservation in the digital world has elevated preservation to be a major function of data management in a way never seen before in libraries. While many have focused on the requirements of the digital content and have fretted over fitting out computer labs and teaching people long out of school how to deal with file formats and mass storage, it is the "curation" in digital curation that is special and is the way forward.

Conclusions

Education for digital curation is very different today than it was ten years ago when the IDCC made its debut. Graduate-level programs are springing up although no one curricular model is as of yet dominant.

Many students graduating from iSchools today are prepared to work with digital content as never before, yet much work lies ahead. Better collaborations between iSchools and digital repositories need to be forged and challenging internships need to be designed. Graduates need an array of skills from data ingest to data visualization to long-term preservation. It is not yet clear if all of this can be covered in 36-hour or even 48-hour programs. What is clear is that the notion of lifecycle curation that involves not only the content but the people involved with the content throughout its useful lifespan is important and needs to be further developed as the foundation of much of the work in cultural institutions.

Educators should not forget their alumni working in the information curatorial professions. There is an intense need for more and better organized and networked continuing education offerings for digital and data curation. Employing institutions also need to support their current staff in their educational efforts and mandate ongoing continuing education.

Funding agencies, and especially the Institute of Museum and Library Services[38] (IMLS) in the United States, the Jisc[39] in the UK, and the European Commission[40] (EC) in Europe have been instrumental in developing digital and data curation programs within today's iSchools. It is doubtful if these programs would have developed as quickly as they have without this support. However, the task is not yet over, as funders could extend their impact by funding projects to explore how to make continuing education more available, attractive, coordinated and feasible for the current workforce while economically sustainable for educational institutions.

Libraries, archives and museums will never ride the crest of the digital wave if they must wait for new employees to bring knowledge of new technologies. The situation is crying out for partnerships between educators and practitioners wherein programs can

38 IMLS: http://www.imls.gov
39 Jisc: http://www.jisc.ac.uk/
40 European Commission: http://ec.europa.eu/index_en.htm

inexpensively be brought to the cultural repository's staff while that institution can provide intensive practicums for today's graduate students.

References

Beagrie, N. (2007). *e-Infrastructure strategy for research: Final report from the OSI Preservation and Curation Working Group*. Retrieved from http://www.nesc.ac.uk/documents/OSI/preservation.pdf

Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science, 323*(5919), 1297–1298. doi:10.1126/science.1170411

Digital Curation Centre. (2005). *Digital curation and preservation: Defining the research agenda for the next decade.* Report of the Warwick Workshop. Retrieved from http://www.dcc.ac.uk/webfm_send/346

DigitalPreservationEurope. (2007). *DPE digital preservation research roadmap*. http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_D72.pdf

Dooley, J., & Luce, K. (2010). *Taking our pulse: The OCLC research survey of special collections and archives.* Retrieved from OCLC website: http://www.oclc.org/content/dam/research/publications/library/2010/2010-11.pdf

Englehardt, C., Strathmann, S., & McCadden, K. (2012). *Training needs survey – Summary report*. Retrieved from http://www.digcur-education.org/eng/Resources/D3.1-Training-needs-survey-summary-report

ISO. (2012a). *ISO 16363:2012. Space data and information transfer systems – Audit and certification of trustworthy digital repositories*

ISO. (2012b). *ISO 14721:2012. Space data and information transfer systems – Open archival information system (OAIS) – Reference model.*

ISO. (2014). *ISO 16919:2014. Space data and information transfer systems – Requirements for bodies providing audit and certification of candidate trustworthy digital repositories.*

Karvelyte, V., Klingaite, N., Kupriene, J., Molloy, L., Snow, K., & Gow A. (2012). *Report on baseline survey and evaluation framework.* Retrieved from DigCurV website: http://www.digcur-education.org/eng/Resources/D2.1.1-Survey-of-existing-training-opportunities

Lee, C. (2009). *Matrix of digital curation knowledge and competencies.* Retrieved from http://www.ils.unc.edu/digccurr/digccurr-matrix.html

Library of Congress. (2010). *Digital preservation survey*. Retrieved from https://www.nedcc.org/eblasts/LOC_DigSurvey.html

National Science Board. (2005). *Long-lived digital data collections enabling research and education in the 21st Century*. Retrieved from http://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf

National Science Foundation. (2006). *To stand the test of time: Long-term stewardship of digital data sets in science and engineering.* A report of the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe, September 2006. Retrieved from http://files.eric.ed.gov/fulltext/ED528649.pdf

National Science Foundation. (2007). *Sustainable digital data preservation and access network partners (DataNet).* Retrieved from http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.pdf

National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. (2003). *Revolutionizing science and engineering through cyberinfrastructure.* Retrieved from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=cise051203

National Science Foundation Cyberinfrastructure Council. (2007). *Cyberinfrastructure vision for 21st century discovery.* Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728_1.pdf

National Science Foundation/Library of Congress Workshop on Research Challenges in Digital Archiving Organizing Committee. (2003). *It's about time: Research challenges in digital archiving and l-term preservation.* Retrieved from http://www.digitalpreservation.gov/documents/about_time2003.pdf

NEDCC Northeast Document Conservation Center. (2015). *Highlights from NEDCC's history.* Retrieved from https://www.nedcc.org/about/history/timeline

Waters, D., & Garrett, J. (1996). *Preserving digital information: Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group.* Retrieved from http://www.clir.org/pubs/reports/pub63