

# Challenges, Curricula, and Competencies: Researcher and Practitioner Perspectives for Informing the Development of a Digital Curation Curriculum

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

## Abstract

*The Digital Curation Curriculum (DigCCurr) project is developing a graduate-level curriculum to prepare students for digital curation in various environments. This paper reports the findings from a survey of digital curation professionals, assessing their perspectives on barriers to digital curation in the field setting; core curriculum competencies and functions; and professional competencies and hiring practices.*

## Introduction

Curation of digital assets is a central challenge confronting libraries, archives, museums, data centers, and other data-intensive organizations. If cultural heritage, science, commerce, health, education and government sectors, among others, are to have long-term access and re-use of meaningful and authentic digital resources, schools of library and information science (LIS) must produce graduates capable of working as digital curation (DC) professionals. Developing curriculum to better prepare today's students to work in this complex arena requires strategies and techniques that address the complete life cycle management of digital objects, from creation to long-term stewardship for future use and access. Presently, limited graduate educational opportunities in DC are available. Only one U.S.-based graduate program, the University of Illinois at Urbana-Champaign, provides a concentration in data curation at the master's level [1].

The School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill (UNC-CH) is developing a graduate-level, openly accessible DC curricular framework, course modules, and experiential components. This project, a collaboration between SILS and the U.S. National Archives and Records Administration (NARA), is made possible by a grant from the Institute of Museum and Library Services (IMLS), titled "Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum," referred to in this paper as the Digital Curation Curriculum (DigCCurr) project [2]. Several DigCCurr data collection activities are on-going to inform the development of a comprehensive, graduate-level DC curriculum, including an extensive review of existing DC literature; a series of 17 semi-structured interviews with international leaders in DC, an analysis of LIS-syllabi of graduate DC-related course components; and an analysis of contemporary job postings for DC-related professional vacancies. These activities have informed the development of a six-dimensional matrix of DC knowledge and competencies, and a 28-point, high-level categorization of DC functions that underlie the curricular framework [3-5]. This paper reports preliminary findings from a fourth data collection activity – a survey administered to

researchers, educators, and practitioners from libraries, archives, academic institutions, data repositories, information centers, cultural heritage institutions, and other organizations working in the areas of DC and digital preservation.

## Methodology

To evaluate and further inform the emerging DC curriculum framework and other project activities, the DigCCurr project team surveyed professionals working in the area of DC. The survey was conducted in two-stages. First, a paper-based survey composed of five open-ended questions was administered April 18-20, 2007 to all DigCCurr2007 symposium attendees [6]. The instrument surveyed attendees' perceptions in three areas: DC challenges at their home institutions; necessary DC curriculum components; and essential DC professional competencies. Content analysis was performed to identify key issues in respondent perceptions. These findings were used to develop an elaborated data collection instrument, resulting in a 25-item, web-based survey consisting primarily of closed-ended questions, with open-ended questions utilized to solicit additional feedback. This paper reports findings resulting from the latter survey.

We emailed survey invitations to 224 DigCCurr2007 symposium attendees in March 2008. Three attendee-supplied email addresses were rejected. Two reminder emails were sent at one-week intervals to the revised sample of 221 attendees. A total of 57 surveys were returned, with two incomplete responses excluded because less than 12% of questions were answered. A total of 55 completed surveys were accepted for analysis, resulting in a 25% response rate. The data was analyzed using both quantitative and qualitative techniques.

## Findings

### Respondent Profile

Data was collected on respondents' employment, including position, location, and supervisory capacity. The majority work in libraries (43%), academic institutions (42%), and archives (32%). Respondents were allowed to choose as many settings as necessary to best describe their institutions. Other settings reported included: research institutes and centers (17%); information and technology services (15%); and others (17%), including museums (4%) and government organizations (8%). Job titles collected represent a diverse field of administrators, practitioners, researchers, and educators working in the areas of DC, digital preservation, and archives and records management. Twenty percent identified themselves as directors, assistant directors, or deputy directors. Other job types reported included: coordinators and departmental heads (17%); librarians (13%); archivists (9%); and managers of

special projects (9%). Asked to report how long they had been in their current positions, respondents offered a time range from four months to 24 years. The average time reported is 6.16 years.

### Local Digital Curation Challenges

At the outset of the survey, respondents were provided with the following definition of DC to provide context for their responses: Digital curation is defined as “the active management and preservation of digital resources over their full life cycle.” Respondents were next presented with seven factors identified in our research as key barriers to implementing DC activities, and asked to indicate how much they thought each factor was a barrier in their institutions. Nearly all (96%) agreed or strongly agreed that the “need to better incorporate digital curation considerations into the institution’s organizational structure, protocols or policies,” was a barrier to implementation, followed by “insufficient communication and coordination between different groups of stakeholders” (91%), and the “need to better identify and operationalize the necessary functions and skills” (89%). The results show some consistency in perceptions of barriers to DC across respondents who, as noted earlier, represent a diverse set of information organizations. It is telling that the barriers appear to be establishment-centered (e.g., policies, communications) as opposed to human-resource centered (e.g., staff) or problem-centered (e.g., scoping problem space). The results appear in Table 1.

**Table 1: Perception of Barriers to Digital Curation**

Barrier	Agree %	Neutral %	Disagree %
Accommodate inst.'s structure/ protocols/ policies	96.3	0	3.7
Improve communication and coordination	90.7	3.7	5.6
Operationalize practices/solutions	88.9	7.4	3.7
Identify and scope problem space	85.2	5.6	9.3
Perceptions of key stakeholders	72.2	11.1	16.7
Lack of staff: count	64.9	20.4	11.1
Lack of staff: expertise	61.1	18.5	16.7

Further information was collected on respondents’ perceived urgency to address these barriers at their institutions, and their capabilities for doing so. Respondents were asked to rank the seven factors based on perceived urgencies and capabilities. The need to improve communications and coordination was ranked as the foremost or second most urgent need to address (45%), followed by the need to change perceptions of key stakeholders (44%) and the need to accommodate institutional structures, policies, and procedures (39%). See Table 2 for rankings of urgencies, grouped by rank assignment.

In comparison, 55% perceived the barrier of identifying and scoping the problem space as the factor their institution had the foremost or second most capability to address, followed by

capability to change perceptions of key stakeholders (37%). The factor for which respondents’ perceived their institutions having the least or second least capability to address was lack of staff assigned (47%), followed by capability to improve communications and coordination (28%). Preliminary analysis shows that while there is some affirmative mapping between urgencies and capabilities – for example, the need to change perceptions of key stakeholders is perceived as both an urgent need and a need institutions’ have the capability to address – there is also a disconnect between urgencies and capabilities, as is the case with the need to improve communications and coordination. See Table 3 for rankings of capabilities.

**Table 2: Aggregated Rankings of Urgency**

Barrier	Most (1 -2) %	Middle (3-5) %	Least (6-7) %
Accommodate inst.'s structure/ protocols/ policies	38.9	51.9	9.3
Improve communication and coordination	45.3	50.9	3.8
Operationalize practices/solutions	22.2	59.3	9.3
Identify and scope problem space	32.1	32.1	32.1
Perceptions of key stakeholders	43.8	27.1	22.9
Lack of staff: count	15.1	35.9	39.6
Lack of staff: expertise	9.4	34.0	47.2

**Table 3: Aggregated Rankings of Capabilities**

Barrier	Most (1 -2) %	Middle (3-5) %	Least (6-7) %
Accommodate inst.'s structure/ protocols/ policies	14.8	61.1	24.1
Improve communication and coordination	29.6	42.6	27.8
Operationalize practices/solutions	24.1	53.7	22.2
Identify and scope problem space	54.7	37.7	7.5
Perceptions of key stakeholders	36.7	36.7	22.4
Lack of staff: count	20.8	22.6	47.2
Lack of staff: expertise	24.5	45.3	22.6

## Curriculum: Core Knowledge and Competencies

DC, by definition, necessitates a broad range of challenges, skills, values, and approaches for the active management and preservation of digital resources across their full life cycle. Respondents were presented with five core categories of DC knowledge and competencies. These components were derived from earlier data collection techniques, and used in this survey to solicit perceptions on desirability of these core competency areas from a broad audience of professionals working in the area of DC.

### *Stages in the Digital Life Cycle*

Respondents were asked to comment on the extent (heavily, moderately, somewhat, or not at all) that the core DC curriculum should emphasize stages in the digital life cycle. A framework of the digital life cycle was provided, comprised of seven stages: 1) design and planning before digital objects are created; 2) digital object creation; 3) digital objects' "life" within the primary use environment; 4) transfer to preservation environment; 5) digital objects' "life" within preservation environment; 6) generation and transfer of copies of digital objects into secondary use environment; and 7) digital objects' "life" within secondary use environment. Overall, across all seven stages of the digital life cycle, the majority of respondents supported either moderate or heavy emphasis of digital life cycle stages in the core curriculum. Specifically, over six out of ten respondents believe that emphasis should be heavily placed on three particular stages: 1) pre-creation (63%), 2) transfer to preservation environment, such as an archival repository (63%), and 3) throughout digital objects' "life" within a preservation environment (64%).

### *Types of Digital Resources*

Respondents were asked to comment on the extent that the core curriculum should emphasize types of digital storage and types of digital objects. Overall, 96% either moderately or heavily support emphasis of digital object types, with 74% reporting heavy emphasis. Respondents were less inclined to emphasize types of digital storage, with 32% supporting heavy emphasis of digital storage types and 34% supporting moderate emphasis. Respondents were asked to provide open-text responses on specific examples of digital objects and digital storage media. While dozens of different references were made to file formats, and hardware and software attributes, there were several calls in the responses to not approach this core curriculum knowledge from a container or system view, but from a conceptual orientation, especially due to the ever-evolving nature of digital objects and storage types.

### *Terminology and Technical Knowledge*

Respondents were asked to comment on the extent that the core curriculum should emphasize specialized terminologies and characteristics of information and communication technologies (ICTs). Overall, support for both categories was similarly affirmative. Ninety-three percent place moderate or heavy emphasis on ICT characteristics, with 39% supporting that these characteristics be heavily emphasized. Respondents were only slightly less inclined to moderately or heavily emphasize specialized terminologies, reported by 83%, with 32% supporting heavy emphasis. Respondents were asked to provide examples of specific types of terminologies. Many referenced the Open

Archival Information System (OASIS) Reference Model, in addition to standards and their associated terminologies. Several respondents also emphasized the need for basic and shared understanding of terminologies to facilitate communications between different discipline-specific languages of professionals engaged in DC, including computer science, archival science, and information and/or library science.

### *Professional Mandates, Values, and Principles*

Respondents were asked to comment on the extent that the core curriculum should emphasize professional ethics, legal requirements, and standards. Notable is that respondents' placed heavy emphasis on standards (76%) when compared to ethics (35%) and legal requirements (43%). Twenty-four percent somewhat support some emphasis on professional ethics, rather than moderate or heavy support.

### *Context: Professional, Disciplinary and Organizational*

In order to identify, develop and evaluate appropriate DC strategies, individuals may need to understand the challenges, traditions, opportunities, characteristics, and dynamics of particular types of professional contexts. Respondents were asked to consider the extent that the core curriculum should emphasize contextual understanding, represented in six response categories: 1) history of professions related to DC; 2) current status and relationships of professions related to DC; 3) trends and potential future directions of professions related to DC; 4) opportunities and strategies for professional development after graduation; 5) disciplinary contexts (e.g., differences between managing digital resources from the social sciences, humanities or physical sciences); and 6) institutional/organizational contexts (e.g., differences between academic, corporate, nonprofit or government environments). No respondents' supported heavy emphasis being placed on history of professions related to digital curation, which is the only occurrence, across all curriculum elements provided in the survey, where no one selected the answer option for "heavily." See Table 4.

**Table 4: Emphasis on Contextualizing Curriculum**

Context	Heavily %	Moderately %	Some-what %	Not at All %
History of professions	0	31.5	64.8	3.7
Related Professions	17.0	49.1	32.1	1.9
Professional development	37.0	37.0	24.1	1.9
Disciplinary contexts	40.7	42.6	16.7	0
Inst. /Org. contexts	29.6	50.0	20.4	0

**Table 5: Emphasis on Digital Curation Core Functions**

Function	Heavily %	Moderately %	Some- what %	Not at All %
<b>TECHNICAL INFRASTRUCTURE</b>				
Systems engineering and development	22.2	42.6	31.5	4.0
Archival storage	37.0	50.0	13.0	0
<b>TREATMENT OF INFORMATION RESOURCES</b>				
Production	26.4	56.6	17.0	0
Transfer	37.7	49.1	11.3	1.9
Ingest	50.0	42.6	7.4	0
Description/Org./Intellectual control	63.0	33.3	3.7	0
Access	50.0	42.6	7.4	0
Use/Reuse/Adding value	48.1	37.0	14.9	0
Destruction/Removal	11.1	55.6	31.5	2.0
Analysis and characterization	58.5	37.7	3.8	0
Validation and quality control	59.3	38.9	2.0	0
Transformation	41.0	51.9	7.4	0
<b>TREATMENT OF METADATA</b>				
Identifying/Locating/Harvesting	35.2	55.6	9.3	0
Data mgmt.	61.1	37.0	2.0	0
<b>HUMAN INTERACTIONS</b>				
Reference/User support	24.5	41.5	28.3	3.8
Coordination w/ external actors	42.6	44.4	13.0	0
Advocacy/Outreach	31.5	48.1	16.7	3.7
<b>STRATEGIES, PRIORITIES &amp; JUDGMENTS</b>				
Selection/Appraisal	46.3	46.3	7.4	0
Preservation: Plan & implement	74.0	26.0	0	0
Analysis/Evaluation of producer environ.	37.7	54.7	7.5	0
Establishing rules, standards & policies	61.1	35.2	3.7	0
<b>ADMINISTRATIVE ACTIVITIES</b>				
Assigning responsibilities	20.4	42.6	33.3	3.7
Budgeting/Resource allocation	24.5	39.6	35.8	0
Mgmt. system configuration	15.1	52.8	28.3	3.8
Planning	61.1	27.8	11.1	0
Project management	51.9	38.9	7.4	2.0
Purchasing/Licensing	9.4	34.0	49.1	7.5
Risk management	35.2	57.4	7.4	0
Security	27.8	59.3	13.0	0

## Curriculum: Core Functions

Next, respondents were presented with six, high-level functional categories derived from earlier research: 1) technical infrastructure; 2) treatment of information resources; 3) treatment of metadata; 4) human interactions; 5) strategies, priorities, and judgments; and 6) administrative activities. First-level sub-functions were presented within each of these high-level categories. Respondents were asked to comment on the extent (heavily, moderately, somewhat, or not at all) that the core DC curriculum should emphasize these sub-functions. Across all categories, the preservation planning and implementation sub-function was designated as deserving the greatest, or heaviest, emphasis, selected by 74%, followed by the description, organization, and intellectual control sub-function (63%). See Table 5 for complete results.

Additionally, respondents were presented with four meta-level functions that may be applied across any of the categories and/or sub-functions presented above: 1) analysis and documentation of curation functions; 2) evaluation and audit of curation functions; 3) research and development to support curation functions; and 4) education and sharing of expertise or guidance on curation functions. Respondents commented on the extent that these meta-level functions should be emphasized. The evaluation and audit meta-function was designated as deserving the greatest, or heaviest, emphasis by respondents, selected by 48%, followed by analysis and documentation (41%). Respondents did not identify any of these meta-level functions as undeserving of any emphasis, though, overall, respondents' placed the least emphasis on research and development. This may reflect the urgency on the part of respondents and their institutions to meet immediate digital curatorial challenges in the practice setting than to explore and plan for responses to future needs or improve on previous responses to past challenges. See Table 6.

**Table 6: Emphasis on Meta-Level Curation Functions**

Function	Heavily %	Moderately %	Some- what %	Not at All %
Analysis/Documentation	40.7	50.0	9.3	0
Evaluation/Audit	48.1	40.7	11.1	0
Research/Development	31.5	48.1	20.4	0
Education	30.0	61.1	9.3	0

## Professional Competencies and Opportunities

Since the intent of a DC curriculum is to prepare students to work professionally in that area, it is necessary to understand the current landscape for professional opportunities in the field. The survey collected information on hiring activities at respondent institutions' for DC-related job vacancies, and identification of desirable professional competencies. Eighty-three percent are involved in hiring activities at their institutions. For those describing their role (n=37), 14% were responsible for making the final hiring decision, while the majority (57%) were part of a group charged with the hiring decision. Regarding employment opportunities, 16% reported posting a digital curation job vacancy within the past month; 27% within the past six months, and 16%

within the past year. Nineteen percent reported posted vacancies more than one year ago, and 12% reported no knowledge of posted vacancies at their institutions.

Further information was collected on respondents' consideration of a set of six professional competencies and characteristics when evaluating candidates for vacancies: 1) personal attributes; 2) interpersonal skills; 3) technical competencies; 4) administrative and managerial competencies; 5) core LIS coursework; and 6) practical experience. Respondents were asked to rank the six attributes, with 1 being most important, 2 second most important, and so on, based on how strongly they were considered when evaluating candidates. For those responding (n=34), practical experience (32%) and technical competencies (29%) were selected as the most important attributes considered. Administrative and managerial competencies was overwhelmingly ranked 6th, selected as the least important attribute considered (41%), followed by core LIS coursework (24%). See Table 7 for rankings of attributes, grouped by rank assignment.

**Table 7: Aggregated Rankings of Professional Attributes**

Attribute	Most (1 -2) %	Middle (3-4) %	Least (5-6) %
Technical competencies	58.8	17.6	20.6
Practical experience	51.5	36.4	12.2
Interpersonal skills	35.3	50.0	14.7
Core LIS coursework	30.3	21.2	45.5
Personal attributes	21.9	50.0	28.1
Administrative and managerial competencies	17.6	23.5	52.9

## Conclusions

Digital curation stands as a grand challenge not only in response to the diversity of tasks and obstacles, both pending and imminent, but also in response to the need to prepare future professionals to meet these challenges. The emerging core curricular knowledge competencies and functional areas identified in this paper have been informed through continuous and extensive outreach to the digital curation community at-large, including leading experts in the field and an ever-growing body of digital curation professionals. Overall, the findings presented here show general affirmation for our developing curriculum. For example, when averaging across the twenty-nine sub-functions for all respondents, four out of ten support heavy emphasis on these functions for the core curriculum, with another four out of ten designating moderate emphasis. Results from this study provide further insight for mapping the needs of the practice setting to deficiencies in the classroom setting, with an intended, end result of an openly accessible, scaleable digital curation curriculum.

## Acknowledgment

This work was supported through IMLS Grant #RE-05-06-0044. Thanks to all the partners and advisory board members of the DigCCurr project. Thanks especially to John C. Schaefer for his invaluable service to the DigCCurr project, particularly the planning and implementation of DigCCurr2007.

## References

- [1] University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science (GSLIS), Master of Science-- Concentration in Data Curation.. [http://www.lis.uiuc.edu/programs/ms/data\\_curation.html](http://www.lis.uiuc.edu/programs/ms/data_curation.html)
- [2] Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum. <http://www.ils.unc.edu/digccurr/>
- [3] Christopher A. Lee, Helen R. Tibbo, and John C. Schaefer, "Defining What Digital Curators Do and What They Need to Know: The DigCCurr Project," in ACM IEEE Joint Conference on Digital Libraries 2007, pp. 49-50. ACM, New York, 2007.
- [4] Christopher A. Lee, Helen R. Tibbo, and John C. Schaefer, "DigCCurr: Building an International Digital Curation Curriculum & the Carolina Digital Curation Fellowship Program," in IS&T Archiving 2007, pp. 105-09. IS&T, Springfield, VA, 2007.
- [5] Christopher A. Lee, Carolyn Hank, and Helen R. Tibbo, "DigCCurr Progress Report: Development of a Graduate-level Digital Curation Curriculum," presented at ALISE 2008, Philadelphia, PA, January 8-11, 2008.
- [6] DigCCurr 2007: An International Symposium in Digital Curation. <http://www.ils.unc.edu/digccurr2007/>

## Author Biography

*Helen R. Tibbo is a Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. She teaches in the areas of archival management and digital curation. She is the co-primary investigator for the NHPRC Electronic Records Research Fellowships Program, and principle investigator for the "Preserving Access to Our Digital Future: Building an International Digital Curation."*

*Carolyn Hank is a Triangle Research Libraries Network (TRLN) Doctoral Fellow at the University of North Carolina at Chapel Hill. She is the project manager for the Carolina Digital Library and Archive's institutional repository. She served as project investigator on the study, "Building from Trust: Using the RLG/NARA Audit Checklist for Institutional Repository Planning and Deployment," and was principle investigator for the recent study, "Blogger Perceptions on Digital Preservation."*

*Christopher (Cal) Lee is Assistant Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill, where he teaches classes in archives, records management and collection development. His research interests include digital preservation, electronic records management, and standardization. He is helping to lead an effort to design and build a long-term institutional repository at UNC. He has an MSI and PhD from the University of Michigan School of Information.*