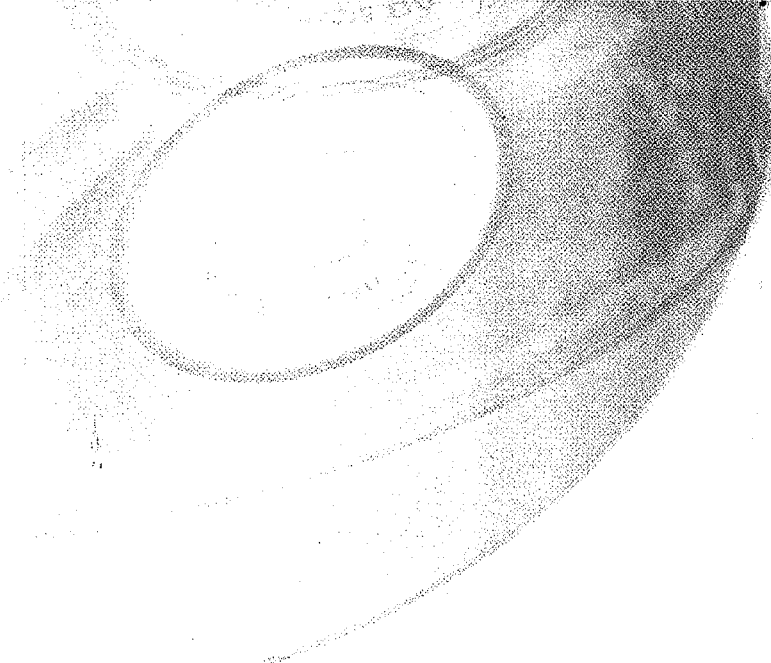




Archiving 2007



Sponsored by the
Society for Imaging Science and Technology

In cooperation with

ALA ALCTS Assoc. for Library Collections and Technical Services
CNI Coalition for Networked Information
DLF Digital Library Federation
DPC Digital Preservation Coalition
ECPA European Commission on Preservation and Access
ISCC Inter-Society Color Council
MCN Museum Computer Network
OCLC Online Computer Library Center
SAA Society of American Archivists

Archiving 2007
Program
and
Exhibits

General Chairs:

Scott A. Stovall, US Government Printing Office

DigCCurr: Building an International Digital Curation Curriculum & the Carolina Digital Curation Fellowship Program

Christopher A. Lee, Helen R. Tibbo and John C. Schaefer; University of North Carolina; Chapel Hill, NC

Abstract

The DigCCurr (Digital Curation Curriculum) project is developing a graduate-level curricular framework, course modules, and experiential components to prepare students for digital curation in various environments. This paper summarizes an initial draft and guiding principles behind a matrix of digital curation knowledge and competencies, which are acting as the basis for our curriculum design efforts.

Introduction

In much of the developed world, digital information is becoming increasingly interwoven into daily life. The volume of digital data being created, transferred, and used is enormous. [1] A recent news item the popular press estimated that "161 exabytes of digital data — or about 161 billion GB — were generated in 2006. And the amount is expected to rise fast." [2] Other estimates of digital data output provide evidence of the enormity of data generated. While becoming dependent on digital tools and the data we produce and consume from them, few individuals or organizations are adequately prepared to take on the long-term care of digital assets.

Long-term access to and use of meaningful, reliable, and authentic electronic records and digital objects is essential to the future of science, commerce, government, education, and cultural heritage. [3-6] Software and hardware obsolescence, media fragility, benign neglect, the expense of metadata creation, and intellectual property rights place many of these materials at risk. Facilitation of on-going access will require concerted, appropriate digital curation efforts. Peter Lyman and Brewster Kahle note that often "the pace of technical change makes digital information disappear before we realize the importance of preserving it." [7]

A decade of work in digital preservation and access has resulted in an emerging and complex life-cycle constellation of strategies, technological approaches, and activities now termed "digital curation," [8-9] but relatively few new educational programs to prepare individuals to work as curators of trustworthy digital repositories. To date, most education and training opportunities focusing on digital preservation and, more recently, curation, have taken the form of one- to five- day workshops. [see, for example, 10-16] A small number of graduate level digital curation programs are emerging including those at the University of Glasgow and the University of Illinois, Urbana-Champaign. [17-18]

The DigCCurr project [19], funded by the Institute of Museum and Library Services (IMLS), is a collaboration of the School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill (UNC-CH) and the

U.S. National Archives and Records Administration (NARA). We are developing a graduate-level curricular framework, course modules, and experiential components to prepare students for digital curation in a wide variety of organizations and with an ever-increasing and changing corpus of digital materials and data formats. DigCCurr brings together an Advisory Board of experts from Australia, Canada, Italy, the Netherlands, New Zealand, the United Kingdom and the United States. The grant includes funding for a set of Digital Curation Fellows who will pursue degrees at SILS beginning in fall 2007. Repositories at UNC will provide practical experience opportunities for the Fellows. Two international symposia will bring the issues of digital curation and this curriculum to the attention of librarians, archivists, museum professionals, data curators, scholars, and the general public.

Project Design and Methods

DigCCurr is a three-year project, which began in July 2006. We are collecting a diverse set of data, which we are analyzing using qualitative analysis software. Our research design is iterative, with numerous opportunities to gain feedback on and revise both our findings and curriculum materials. Our approach is to blend models, guidance documents and theoretical frameworks concerning digital curation and preservation with insights from practitioners who are creating curation processes in their repositories. By exploring what the first generation of digital curators are doing and discussing with digital preservation researchers what challenges lay ahead, we will derive a proposed body of knowledge, skills, and perspectives that educational programs should provide to their students. We recognize that a digital curation curriculum can never be considered finalized but must instead change in relationship to the world it serves.

At the start of our project (July 16-21), Cal Lee, John Schaefer, and Helen Tibbo, participated in the workshop entitled, *Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems*, run by the Cornell University. [12c] We gained many insights to inform our curriculum development, and the binder of supporting materials has also served as a valuable resource that reflected the state-of-the-art of continuing education offerings for digital preservation as of the summer of 2006.

We are obtaining and analyzing a variety of documentary sources. We have carried out, and continue to supplement, a detailed review of existing literature relevant to digital curation and digital curation education. Within the scope of our data collection are also syllabi from courses and workshops (at UNC and elsewhere), job advertisements for positions related to digital curation, workshop and other continuing education

opportunities in digital curation, and relevant standards, tools, services, and research projects.

The 17 members of our Advisory Board are serving as valuable sources of data and guidance. We have conducted semi-structured interviews with all Board members, gaining their insights on the following issues:

- ∞ How to define "digital curation"
- ∞ The main functions or activities necessary for digital curation
- ∞ Main topics to cover in a digital curation curriculum
- ∞ Distinct roles that are needed in digital curation (and associated curriculum requirements)
- ∞ Sorts of environments in which graduates of a digital curation program should be able to work (and associated curriculum requirements)
- ∞ What they would look for in a good candidate, if hiring for a digital curation job
- ∞ Important skills or knowledge required for digital curation that may be lacking within the professions engaged in the work (and associated curriculum requirements)
- ∞ Characteristics of a good practical field experience

We plan to conduct an online survey with our Board members in order to clarify, refine and identify priorities within our matrix of digital curation knowledge and competencies (see below).

On April 18-20, 2007, we will host an international symposium on digital curation in Chapel Hill.[20] The program features panel discussions and presentations by internationally recognized experts from a broad range of academic and national libraries, data centers, information schools, corporations, and government entities. The symposium will provide a public forum and visibility for digital curation issues. It will also serve as a valuable source of data for the DigCCurr project. In addition to coding and analyzing the presentations and papers, we will conduct two sets of focus group discussions during the symposium. After the symposium, we will meet with members of the Advisory Board, sharing and soliciting comments on our products to that point.

Throughout the course of the DigCCurr project, we will be disseminating and gaining feedback on project deliverables at a diversity of professional conferences. We will also host a second digital curation symposium in Chapel Hill in Fall 2008, which will allow us to share our results, advance a community of practice surrounding digital curation, and discuss a professional agenda for next steps. Sustainability and updating of the educational products will be supported through programmatic adoption at UNC and by encouraging widespread dissemination and use at other educational institutions worldwide.

Matrix of Digital Curation Knowledge and Competencies

Based on a review of existing literature, other documentary materials, and semi-structured interviews with the members of our Advisory Board, we have developed a 6-dimensional matrix for identifying and organizing the material to be covered in a

digital curation curriculum. Each unit of curriculum content will address one or more dimensions. The six dimensions are:

Fig. 1: Matrix for Curriculum Development

1. Type of Resource • Level of Aggregation • Level of Abstraction • Medium • Format • Genre
2. Functions and Skills [see below]
3. Professional, Disciplinary or Institutional/Organizational Context • Professional Context • Disciplinary Context • Institutional/Organizational Context
4. Mandates, Values and Principles • Ethics • Legal Requirements • Standards • Interoperability and Sustainability Requirements
5. Prerequisite Knowledge • Terminology • Characteristics of Technologies
6. Lifecycle Stage • Pre-Creation Design and Planning • Creation • Primary Use Environment (Active Use) • Transfer to Archives • Archives (Preservation Environment) • Transfer Copies or Surrogates to Secondary Use Environment • Secondary Use Environment

A curriculum unit can focus on a dimension in general or specifically as it intersects with one or more other dimensions. For example, one could teach a general unit on digital preservation (main considerations and practices), but one might also want to teach a unit specifically on preservation of video, images, or text; during the active use stage or within the archival environment; within a corporate recordkeeping context, or within a collecting repository environment or some combination thereof.

Guiding Principles

Our design of both the matrix and curriculum units is guided by a set of basic principles.

Principle 1: Build on an installed based

We are drawing from both existing and emerging models, frameworks, guidance documents and educational initiatives. Starting with these educational materials we are seeking input from the professional curation community to anchor our program in the reality of the work place. There are many streams of professional activity that can feed into a strong digital curation curriculum. Additionally, we are monitoring and drawing from on-going research projects worldwide to include the

challenges curators face now and in the future. The DigCCurr2007 symposium in April and the intensive guidance from our Advisory Board members will both provide extremely valuable information on recent initiatives and likely future trends.

Principle 2: Digital curation activities span the entire lifecycle of digital resources

Responsible digital curation is much more than preservation of bits. It reaches into both the creator/producer and user environments associated with collections of digital objects. The foundational vision of the Digital Curation Centre (DCC) in the UK is that “long term stewardship of digital assets is the responsibility of everyone in the digital information value chain” and that “the maintenance, usability and survival of digital resources depends on regular planned interventions; care needs to be taken at conception, at creation, during use, and as use transitions to lower levels” [9]. As discussed in the JISC e-Science Curation Report (2003), digital curation, spanning the lifecycle of digital materials, is broader than the function of digital archiving. Digital preservation is, in turn, a component of digital archiving. [21] Several members of our Advisory Board have suggested that one of the major benefits of adopting the label “digital curation” is that it foregrounds the importance of issues and activities that fall outside the scope often associated with “digital preservation.”

Principle 3: Keep lifecycle stages simple, and move complexity into the functions

We have adopted this principle for two major reasons. First, this allows us to map our lifecycle stages to existing domain-specific lifecycle models. Second, many activities that are important to digital curation (e.g. classification, selection, metadata extraction, characterization, file format transformations) can occur at more than one stage of the information lifecycle. For example, many of the activities that the Reference Model for an Open Archival Information System (OAIS) [22] places within the Ingest function can be important and valuable to carry out, not only during transfer to an archives, but also during system design, creation, active use, within the preservation environment, during transfer to a secondary use environment and within the secondary use environment.

We have identified and analyzed numerous lifecycle models, which often include considerable detail about the processes and workflow of a particular domain or context. From these models we have identified a limited, high-level set of lifecycle stages: Pre-Creation Design and Planning, Creation, Primary Use Environment (Active Use), Transfer to Archives, Archives (Preservation Environment), Transfer Copies or Surrogates to Secondary Use Environment, and Secondary Use Environment. (See Fig. 1.)

Principle 4: Build from modules, rather than entire courses

We have identified several categories of digital curation knowledge and competency that should be addressed within a

curriculum, which either do not warrant an entire course or may not fit well into existing courses or areas of instructor competence. Two such categories that have emerged from our analysis are legal requirements (e.g. compliance with Sarbanes-Oxley, Freedom of Information Act, intellectual property rights, and copyright) and understanding technologies (e.g. basic data structures, operating systems, storage environments). We are exploring approaches for integrating small modules into the curriculum that address such particular topics. This can be a challenge in many university environments, in which the 3- or 4-credit-hour course is the basic unit for administering curricula. We are identifying and developing relationships with an international set of individuals who have expertise in areas represented by the modules of the digital curation curriculum. We are exploring options for the experts to provide instruction through a combination of physical visits and the use of remote instructional technology.

Principle 5: Emphasize core, generalizable modules

If each of the six dimensions in the matrix had ten categories of values (rows), there would be one million individual cells. Obviously, no curriculum could devote attention to all of the cells individually. The aim should be to teach units that address a large number of cells simultaneously, while also supporting a drilling down into specific cells, when necessary (both through coursework and practical field experiences).

Functions and Skills

During the first phase of DigCCurr, we have devoted considerable attention to the development of a taxonomy of functions and skills. This is an iterative process; we expect that the taxonomy will undergo significant revision based on further analysis of received comments and collected literature, syllabi, and job postings. Figure 2 characterizes the current draft of our taxonomy.

Fig. 2. Taxonomy of Digital Curation Functions and Skills

First-Level Functions and Skills:
Access
Administration
Advocacy and Outreach
Analysis and Characterization of Digital Objects/Packages
Analysis and Evaluation of Producer Information
Environments
Archival Storage
Collaboration, Coordination and Contracting with External Actors
Common Services
Data Management
Description, Organization and Intellectual Control
Destruction and Removal
Identifying, Locating and Harvest
Ingest
Management
Preservation Planning and Implementation

Production
 Purchasing and Licensing of Content
 Reference and User Support Services
 Removal from Archive
 Selection, Appraisal, and Disposition
 Systems Engineering and Development
 Transfer
 Transformation of Digital Objects/Packages
 Use, Reuse and Adding Value by User
 Validation and Quality Control of Digital Objects/Packages

Meta-Level Functions and Skills:

Analysis and Documentation of Curation Functions
 Evaluation and Audit of Curation Functions
 Research and Development to Support Curation Functions

Digital Curation Fellows and Practical Field Experiences

The DigCCurr grant is funding five Digital Curation Fellows, who will build on what they learn in the curriculum by participating in practical field experiences designed by our campus project partners. [23] The field experiences will provide Fellows with the opportunity to contribute to the management of a wide range of digital objects including public records, cultural heritage assets, teaching materials, and research data. These Fellowships will serve to integrate the curriculum and the experiential components, advertise the existence of the programs at SILS, draw attention to the need for digital curation, and provide an essential empirical testing ground for the viability and appropriateness of the curriculum content in specific contexts.

Integration into Curriculum at SILS

In addition to developing products to be used by digital curation educators and professionals more broadly, the DigCCurr project is also developing and setting the groundwork for a digital curation curriculum at the UNC School of Information and Library Science. As a test bed for the curriculum, we will confront a number of challenges and opportunities. Our analysis of course syllabi at SILS has identified a large number of valuable elements relevant to digital curation, which are already being taught in our school. We have also been fortunate to have a complementary project -- developing a framework and set of modules for a digital library curriculum -- currently underway at SILS and Virginia Tech. [24] We continue to analyze the existing SILS curriculum to determine opportunities and potential areas of revision or expansion. We are also looking to other course offerings at UNC and nearby institutions. In the coming year, we will be translating our conceptual products into more specific decisions about course expectations for the Digital Curation Fellows and other SILS students interested in pursuing careers in digital curation.

Acknowledgements

This work was supported through IMLS Grant # RE-05-06-0044. Thanks to all the partners and advisory board members of the DigCCurr project. We would also like to

acknowledge Jeanette Bastian and Elizabeth Yakel for generously sharing with us data they collected from syllabi in archival education, which are serving as an extremely valuable resource for our own syllabus analysis.

References

- [1] Peter Lyman and Hal R. Varian, "How Much Information 2003?" University of California -- Berkeley. 2003.
<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>
- [2] Michelle Kessler, "Days of Officially Drowning in Data Almost upon Us." USA Today. March 5, 2007.
http://www.usatoday.com/tech/news/2007-03-05-data_N.htm
- [3] Seamus Ross, "Changing Trains at Wigan: Digital Preservation and the Future of Scholarship." NPO Preservation Guidance Occasional Papers (National Preservation Office, London, 2000).
<http://www.bl.uk/services/preservation/occpaper.pdf>
- [4] Helen R. Tibbo, "On the Nature and Importance of Archiving in the Digital Age." *Advances in Computers*, 57, 1-67 (2003).
- [5] DLM-Forum, Proceedings of the DLM-Forum on Electronic Records: European Citizens and Electronic Information: The Memory of the Information Society. Brussels, October 18-19, 1999. (European Communities, Luxembourg, 2000).
<http://europa.eu.int/ISPO/dlm>
- [6] B. Bergeron, *Dark Ages II: When the Digital Data Die* (Prentice Hall PTR, Upper Saddle River, NJ, 2002).
- [7] Peter Lyman and Brewster Kahle, "Archiving Digital Cultural Artifacts" *D-Lib Magazine* (July/August 1998).
<http://www.dlib.org/dlib/july98/07lyman.html>
- [8] Digital Curation Center, "What Is Digital Curation?"
<http://www.dcc.ac.uk/about/what/>
- [9] Chris Rusbridge, Peter Burnhill, Seamus Ross, Peter Buneman, David Giarretta, Liz Lyon, and Malcolm Atkinson. "The Digital Curation Centre: A Vision for Digital Curation." Paper presented at From Local to Global: Data Interoperability--Challenges and Technologies, Mass Storage and Systems Technology Committee of the IEEE Computer Society, Sardinia, Italy, June 20-24, 2005.
http://www.dcc.ac.uk/docs/DCC_Sardinia_paper_final.pdf
- [10] For continuing education opportunities from the Northeast Document Conservation Center (NEDCC): see "Education" at:
<http://www.nedcc.org/education/introduction.php>. Of particular note has been the School for Scanning.
- [11] Digitization for Cultural Heritage Information Professionals. Week-long workshops held at Humanities Advanced Technology and Information Institute (HATII). University of Glasgow, Rice University, and the University of North Carolina at Chapel Hill, 1998-2004.
- [12] Through its University Library, Cornell University has offered three important workshop series from 1996 through 2006: a) Digital Imaging for Libraries and Archives; b) Moving Theory into Practice: Digital Imaging for Libraries and Archives, <http://www.library.cornell.edu/preservation/workshop/>; and c) Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems.
<http://www.library.cornell.edu/iris/dpworkshop/>
- [13] For ERPANET educational offerings see: erpaEvents -- Seminars and Workshops given June 2002 -- October 2004.
<http://www.erpanet.org/events/index.php>
- [14] For Digital Curation Center (DCC) educational offerings see: DCC Events -- Workshops, Information Days, Conferences, and Lectures & Talks. <http://www.dcc.ac.uk/events/>
- [15] For the Digital Preservation Coalition's (DPC) educational offerings see "Forums and Meetings"
<http://www.dpconline.org/graphics/events/>

- [16] For DigitalPreservationEurope (DPE) educational offerings see <http://www.digitalpreservationeurope.org/>
- [17] Humanities Advanced Technology and Information Institute (HATII). University of Glasgow. MSc in Management and Preservation. <http://www.hatii.arts.gla.ac.uk/imp/index.htm>.
- [18] University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science (GSLIS), Master of Science--Concentration in Data Curation, with fellowships funded by the Institute of Museum and Library Services. http://www.lis.uiuc.edu/programs/ms/data_curation.html
- [19] Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum. <http://www.ils.unc.edu/digccurr/>.
- [20] DigCCurr2007. <http://ils.unc.edu/digccurr2007>
- [21] Philip Lord and Alison Macdonald, e-Science Curation Report. Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision, prepared for: The JISC Committee for the Support of Research (JCSR). (The Digital Archiving Consultancy Ltd., Twickenham, UK, 2003). http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf
- [22] CCSDS, Reference Model for an Open Archival Information System (OAIS): Blue Book CCSDS 650.0-B-1 (2002).
- [23] Our campus partners include The Academic Affairs Library, ibiblio, The Odum Institute, and Information and Technology Services.
- [24] Curriculum Development: Digital Libraries. A collaborative research project with Virginia Polytechnic Institute and State University Department of Computer Science. Curriculum development for programs in Computer Science and Information and Library Science emphasizing digital libraries, consistent with the Computing Curriculum 2001 (CC2001), a joint effort of ACM and IEEE-CS, defining curricula for computer science.

Author Biography

Christopher (Cal) Lee is Assistant Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill, where he teaches classes in archives, records, management and collection development. His research interests include digital preservation, electronic records management and standardization. Lee 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.

Helen Tibbo is an Professor in the School of Information and Library Science at the University of North Carolina - Chapel Hill. She teaches in the areas of archival and records management, and digital curation. Dr. Tibbo is a co-investigator on the NSF-sponsored VidArch project with Gary Marchionini, Cal Lee, and Paul Jones. Dr. Tibbo is also co-principle investigator for the Mellon-sponsored Developing Standardized Metrics Project, and directs the NHPRC Electronic Records Research Fellowships. She is PI for the DigCCurr project.

John C. Schaefer is a doctoral student at the School of Information and Library Science at the University of North Carolina at Chapel Hill. His research interests include digital curation, archeology, reuse of research data, and scholarly communication. He has an MA from East Carolina University and an MSLS from SILS.