Digital Curation for Digital Natives

Elizabeth Yakel, Paul Conway, Margaret Hedstrom and David Wallace

School of Information University of Michigan, 105 South State Street, North Quad 4419, Ann Arbor, MI 48109-1285. E-mail: yakel@umich.edu, pconway@umich.edu, hedstrom@umich.edu, and davwal@umich.edu

Digital information is all around us. More and more information is either born digital or digitally reformatted. A new generation of digital archivists and digital preservation specialists (also known as digital curators) is needed to manage this information throughout its life cycle. A strong digital curation curriculum has three components: courses, practice-based internships, and a solid technological infrastructure. This article discusses the education of digital archivists and curators at the University of Michigan's School of Information in the context of these three components.

Keywords: archival education, digital preservation, digital curation, preservation education, teaching with technology, service learning

Introduction

Digital information is all around us. More and more information is either born digital or digitally reformatted. A new generation of professionals is needed who are comfortable working in hybrid (digital and analog) environments and capable of managing media-neutral information throughout its life cycle. A strong digital curation curriculum in an iSchool environment is one possible response to this need. This article discusses the development of a digital curation curriculum at the University of Michigan's School of Information (SI), which has developed a workable and systematic curriculum over the past three years that contains three interrelated components: courses, internships, and technology tools that provide a holistic education for our students.

We entitled this piece Digital Curation for Digital Natives because we are aware of the conundrums inherent in teaching digital curation principles and practices to the generation born since 1985. This cohort has been variously designated the Net Gen, Google Gen, Digital residents, Millennials, and Digital natives. The so-called "digital natives" differ from previ-

ous generations of students for reasons beyond their general fluency with technology. They have grown up in a world of constant connectivity. They demand more active learning and feedback and expect courses to incorporate more interaction among students in the classroom and beyond (Oblinger, 2003). On the surface, they are at ease with technology. But even though technology is intertwined in our students' lives, many do not possess the information literacy skills or strategies for learning with technology or learning how to learn new technologies (Bennett, Margaryan, & Littlejohn 2008; Sharpe, Beetham, Benfield, DeCicco, & Lessner, 2009). One might think that digital curation would be a subject of natural interest for our students, because pervasive technology depends upon an aura of persistence—always on, always accessible, dependable, reliable, and trustworthy. But the reality is that digital curation is perhaps one of the hardest topics to teach precisely because ubiquitous computing dulls the sense of urgency and reinforces a sense of complacency that only those deeply immersed in the technical challenges of digital curation understand to be a chimera.

Background

Since the late 1960s, the University of Michigan's School of Information (SI) has offered master's-level course work in the administration of archives and records. As general interest in archival issues has increased, the School has increased its commitment to retaining faculty talent and augmenting its course offerings beyond the core constructs of the archival community. Issues such as appraisal and preservation have been covered in increasing depth and breadth through specialized course offerings. In the mid-1990s, when the entire master's curriculum was reformed and the iSchool developed, the archives curriculum was expanded and revised to meet professional demands and to cover digital as well as analog materials. The resulting seven archival courses (Introducing Records and Archives: Principles and Practices, Access Systems for Archival Materials, Appraisal of Archives, Electronic Records Management, Preserving Information, Research Seminar on Archives and Institutions of Social Memory, and the Practical Engagement Workshop in Archives) formed the Archives and Records Management (ARM) specialization, one of the four original specializations at SI along with Human Computer Interaction, Library and Information Services, and a Tailored option. This philosophical approach to introducing digital and analog materials in all of our courses remains a hallmark in our curriculum. Integrative teaching about digital and analog materials is essential, as we envision our graduates working in hybrid environments.

Since the ARM specialization was introduced, we have seen our students placed in increasingly diverse organizations and positions. Many graduates now deal principally with digital information, and more and more professional positions ask for a variety of digital curation skills. Concurrently, the explosion in digital curation research and professional activity made adequate coverage increasingly difficult, and an increasing number of students expressed interest in internships and careers in digital preservation. Therefore, in 2007 we proposed a separate specialization, Preservation of Information (PI), to address digital curation and preservation issues. This enabled us to briefly introduce topics in the ARM courses and then treat them in greater depth and specificity in PI. Students seeking to satisfy the requirements for the PI specialization must complete twelve credits in this area; up to three can be internship credits. With careful planning, students can complete the PI specialization and an additional specialization, typically either ARM or Library and Information Services (LIS).

Courses

In planning for this specialization, we proposed the following nine courses:

- Preserving Information (a required introduction)
- Advanced Preservation Administration
- Digital Preservation
- Web Archiving
- Digitization for Preservation
- Preserving Sound and Motion
- Economics of Sustainable Digital Preservation
- Physical Treatment Processes for Preservation Administrators
- Science/Social Science Data Curation

We viewed this as a comprehensive suite of courses that would introduce students to general concepts (in Digital Preservation), provide more detailed knowledge about certain media (Web Archiving and Preserving Sound and Motion), and introduce students to curation in specific arenas (Science/Social Science Data Curation). Since many of these were new courses, we were able to fashion them to fit together with minimal overlap. We have also designed these courses to cover the managerial, structural, technical, and financial issues associated with digital cu-

ration. All of these courses are now offered in some format. We have added two more courses, Data Manipulation and a Practical Engagement Workshop in Digital Preservation, totaling 11 potential courses. In order to increase flexibility, each course is offered as a half term course (7 weeks). This assists us in balancing faculty teaching loads and enables us to mix and match various courses to ensure that all are offered at least once every two years (the time required to complete the master's degree if pursued full time).

Active learning is achieved through a combination of lecture and small group discussion or exercises. In discussions, we encourage students to reflect on how class concepts might or might not apply in their work or internship settings. We also utilize a course management system that has chat and wiki capabilities. These tools are used in different ways to increase interaction outside of the formal classroom. Within each course, we also balance individual and group assignments. In order to involve the students in real-world problems, many of the assignments concern something they might do in a professional position after graduation (e.g., needs assessment or web crawl).

The content of these courses is wide ranging. Synopses follow:

- *Preserving Information* is the introductory course that provides a broad introduction to the field. This course is a prerequisite for most of the others and is the one required course for the PI specialization. To immediately engage the students in applying classroom skills in the field and then reflecting on this process, the final group project for this course is a preservation needs assessment. In many of these courses, the assignments are constructed to help students transfer classroom knowledge into different contexts and then reflect on the experience.
- *Advanced Preservation Management* continues where Preserving Information

leaves off and covers facility, policy, and managerial issues in greater depth. In this course students build on the needs assessment to create a preservation plan for the same organization.
- *Digital Preservation* covers the core components, functions, and activities of an archival information system capable of acquiring content created outside of the system, managing the information for the long term, and delivering meaningful content to end users. The course is organized around the Open Archival Information Systems Reference Model (OAIS) (Consultative Committee for Space Data Systems, 2002), which is an international standard for digital preservation systems. Students work with a variety of applications (e.g., Checksum, JHOVE) that illustrate the basic concepts of digital preservation and curation; share evaluations of technology, file format, and metadata standards through the course wiki; and perform a TRAC or DRAMBORA audit on a real organization.
- *Web Archiving* familiarizes students with the tools and appropriate techniques for the preservation and curation of information delivered through the "surface" Web (static Web pages, blogs, E-mail discussion lists, etc.) and introduces strategies for acquiring "deep" Web content (e.g. databases, streaming media, and authenticated resources). Students are required to design, execute, and critique a web crawl on a topic of their choice.
- *Digitization for Preservation* concentrates on the standards, techniques, and metadata requirements for the creation of digital surrogates for long-term preservation. The course also touches on why libraries and archives might want to engage in digitization for preservation and on policy setting around the digital representation of artifacts.
- *Preserving Sound and Motion* examines and evaluates the archival field's current preservation standards for storage

and duplication of machine-dependent media, including especially magnetic audio and video tape and successive generations of motion picture film. The course juxtaposes the management of fragile media with options for digitally reformatting.

- *Economics of Sustainable Digital Preservation* covers economic mechanisms and cost models through the lens of different approaches to sustainable digital curation in the United States and internationally.
- *Physical Treatment Processes for Preservation Administrators* introduces students to the materiality of books and papers and discusses the application of those principles to developing policy and working with vendors of conservation and reformatting services.

We were able to take advantage of an Integrative Graduate Education and Traineeship (IGERT) Award from the National Science Foundation to develop a three-credit doctoral-level seminar on Data Curation (http://opendata.si.umich.edu/). The course was designed as a required core course on data curation, but advanced master's students may enroll with the permission of the instructor. This course introduces the foundational principles, requirements, and techniques for data curation. Students learn about the concepts, language, terminology, and tools of data curation, become familiar with the current state of knowledge, and identify research issues in the field. The course is divided into three parts: elements needed for the curation of research data, data repository development, and individual mechanisms and technical elements of a data repository necessary for the management, storage, preservation, discovery, and use of data and the relationships among these factors. At some point, this course will be revised for master's students and become the Data Curation course that was in our original plan for the Preservation of Information specialization. This course is cross-listed with the Department of Electrical Engineering and Computer Science.

Data Manipulation was originally developed to support the Information Retrieval specialization at SI. We included it in the PI specialization because it interfaces well with Web Archiving. Students enrolled in Data Manipulation learn data harvesting, processing, and aggregation of large-scale data sets, such as a web crawl. Specific techniques introduced include automated means of gathering, parsing, and summarizing data. SI offers several other courses that are not part of the PI specialization but are relevant to students interested in digital preservation careers, such as the Seminar on Digital Libraries and Archives and the ARM course, Electronic Records Management.

As a whole, the ARM and PI courses address many of the topical areas listed in various digital curation curricula guidelines (Lee, 2009; Lee, Tibbo, & Schaefer, 2007). These master's courses also support data curation and preservation learning through a series of assignments that build on one another and simulate real-world activities (Table 1).

As noted in the course descriptions, we are employing a variety of technologies in our courses. This emphasis and our strategies for doing so follow in the next section.

Technology

Current digital curation students need to learn how to learn about new tools (both software applications and systems), how best to use the tools as part of a normal problem-solving routine in the management of digital collections, and how these tools can be used to support core archival values, such as provenance and authenticity, and core curation functions, such as preservation. Although we had been using different technologies in our courses for some time, as instructors we were not completely satisfied with how these worked in the classroom setting. We

Table 1: Courses and Assignments.

Course Name	Assignment
Preserving Information	Preservation Assessment
Advanced Preservation Management	Preservation Plan
Digital Preservation	TRAC or DRAMBORA Assessment
Web Archiving	Design and Execute a Crawl
Digitization for Preservation	Scanner and object test targets and analysis software
Preserving Sound and Motion	Digitization project assessment and plan of work
Economics of Sustainable Digital Preservation	Case study of a digital repository focusing on how it manages sustainability
Data Manipulation	Collecting, analyzing, and visualizing a large-scale dataset
Data Curation	Designing a research study around a data curation topic

wanted students not only to be able to use the tools but to evaluate them in terms of how well they support archival and curation processes and workflows. The integration of the tools in class and in the assignments was difficult in the past, and the dual role of professor and systems administrator could be overwhelming. Introducing archival and preservation technologies is also difficult because our digital natives are used to more polished interfaces and self-explanatory applications than those in these areas.

Our initial experimentation with technology in the classroom led us to develop a National Endowment for the Humanities grant, the Preservation and Access Virtual Education Laboratory (PAVEL) (http://pavel.cms.si.umich.edu/). The grant specifies that we integrate a variety of technologies in five separate courses. Table 2 presents a list of the courses and the technologies to be used. We will also publish specifications, course learning objectives, assignments, and lecture notes and content online for others to use.

Receipt of this grant came at an oppor-

tune time, since the School of Information is taking up residence in an innovative living-learning complex that is foregoing fixed computer laboratories. The move to a virtual laboratory environment for technically grounded courses necessitates changes in the way we deliver our applications. At the same time, a change in the pedagogical environment creates an opportunity to establish platform independence (e.g., MAC, PC) for preservation-oriented software systems and to enable students to use the applications in a classroom setting or remotely on campus or at home.

The PAVEL grant enables us to engage in several activities to better integrate technology into courses. There are three aspects to this integration: institutionalization of the tools for sustainability, creation of course materials and assignments, and evaluation to ensure we are meeting our learning objectives and not just teaching to the tool. First, it will help us introduce the curation and access applications to our School's computing staff so that they can provide better systems to support these

Table 2: Technology Used in Preservation Courses.

Course Name	Technology
Web Archiving	ArchiveIT
Digitization for Preservation	GoldenThread (Image Science Associates)
Digital Preservation	FITS, BagIT, Checksum
Records Management	Sharepoint
Access Systems for Archival Materials	Archivist's Toolkit, Archon, ICA AtoM, Collective Access

applications. This will off-load some of the system administration responsibilities from the instructors and, more importantly, help us institutionalize use of these applications. It will also enable the instructors to focus on student learning. Preservation and archival content management system interfaces are very basic. In the words of one of our students, "I remember when I first opened it I was a little freaked out by how few options the menu bar held!" Student use entails not only a learning curve but also a paradigm shift away from applications and interface features with which they are familiar.

Second, we are using the grant as an opportunity to rethink courses and revise the syllabi (course objectives, the outline of the course, assignments, and lectures). In the past, we had essentially added technology into existing courses and made only minimal changes to accommodate a given technology. This was less than optimal, and we were ready to do major course redesign. The grant enables this to be done systematically, as we are also able to redo the infrastructure and evaluation elements of these courses.

Third, the grant has a strong evaluation component that focuses not only on the tools, but also on courses and our course sequencing. We have hired an education evaluator to help us get the most out of our experimentation (with the technology, assignments, etc.) during the grant period. This will also result in evaluation tools that will be invaluable as we continue to introduce and use new technologies in the upcoming years. We envision several types of evaluation: (1) an initial survey to assess students' technological expertise, their archival or digital preservation experience, and their comfort in learning new technologies; (2) evaluation of student work; and (3) course (and tool) evaluation. Our goal is not to teach to the tool, but to use the tools to enable students to learn how to learn new tools, to strengthen their archival or curation conceptual framework (e.g., "thinking like an archivist or digi-

tal curator"), and to use the tools to meet learning objectives, whether those deal with metadata transformation, workflow, or meeting selection objectives in a web crawl.

Table 2 presents a list of the courses and the technologies to be implemented.

Since this article focuses on digital curation, we highlight the selection and use of technologies in two courses closely related to this subject: Digitization for Preservation and Web Archiving. Our goal in all of the courses was to select technologies that strengthened other skills necessary to "think like a digital curator."

For Web Archiving we selected Archive-IT. Students in this course design and execute a web crawl and then perform the initial steps in curation of the content. Students have to understand something about the problem space (the Web) as well as the selection, representation, retrieval, and storage issues that are involved in the crawl. Previously we had used HTTrack, a free, open source website copier and offline browser that worked well for the OAIS ingest functions but required considerable customized effort for data and file management. Furthermore, it had no access and retrieval functionality. Each program has advantages and disadvantages. Archive-IT allowed the students fewer options for setting the parameters of each crawl than HTTrack, but had the advantage of providing comprehensive reports on the crawl sessions and the data gathered in each session. Archive-IT provided standard templates for adding metadata, but it did not allow much flexibility in designing or choosing alternative descriptive frameworks. Neither Archive-IT nor HTTrack supported long-term content management or presentation of archived collections beyond basic web browser functions. As a result, both support different learning objectives and give students a chance to exercise slightly different skill sets.

Digitization for Preservation involves a collaboration with Image Science Associates of Rochester, New York that

gives students direct access to a powerful analytical tool for assessing the quality of digital scanners and learning about the use of standardized test targets in production scanning environments. In an exercise, students seek out operational scanners on campus (both flatbed and overhead cameras), create test scans of a variety of archival and library materials utilizing the ISA test targets, and then process the resulting scans through the GoldenThread software available through Michigan's virtual sites computers. The software produces a variety of diagnostic reports, which the students interpret and assess. The exercise mimics the processes of scanner calibration that are fundamental to preservation-grade digitization, all the while exposing students to the complexities of obtaining color fidelity in today's environment.

As we developed the Preservation of Information specialization, we realized that the courses and technologies were just two of three elements required to educate digital curators. Outside the classroom, internships are needed in order to test and transfer knowledge from the classroom into practice and then reflect on this process.

Internships

One of the philosophical underpinnings of the School of Information's Master of Science in Information (MSI) degree is a belief that students need to engage with communities and apply what they are learning in the classroom. Internships (or Practical Engagements as they are known at SI) are an integral part of educating all master's students. During the 48-hour master's program, all students must earn eight Practical Engagement (PE) credits to graduate. Students receive PE credits by participating in mentored internships as well as by enrolling in some courses that enable students to do projects in the community. All PE internships that are not linked to a specific course require a solid project proposal and an on-site mentor for

the student, both approved by the SI Career Services office. Students receive one credit hour for every 60 hours of internship work completed, so students obtain a minimum of 480 hours of work experience during their time at SI.

As we began building the PI specialization, two things became apparent. First, SI students had few opportunities to undertake digital preservation and curation internships, either locally or in the wider community of practice. Therefore, we needed to recruit institutions and mentors specifically in this area of practice. Second, while Ann Arbor was home to several institutions with stellar digital preservation and curation programs, such as the University of Michigan Libraries and the Inter-university Consortium for Political and Social Research (ICPSR), many more units in the university and organizations in the community confronted data curation issues but were unable to obtain student interns without some outside help. To fill these gaps, the Institute for Museum and Library Services provided funding for the "Engaging Communities to Foster Internships for Preservation and Digital Curation" project (http://preservation. cms.si.umich.edu/). Through the grant, the School of Information established ten summer internships at institutions with strong digital curation and preservation programs. Our summer intern partners are the Center for Research Libraries, the Florida Center for Library Automation, the LOCKSS program at Stanford University, the Northeast Document Conservation Center, OCLC, the Smithsonian Institution Archives, Safe Sound Archive, the Inter-university Consortium for Political and Social Research, the University of Michigan Libraries, and the Internet Archive. Each of these institutions agreed to host one intern per year for a three-year commitment to provide 360 hours of work during the summer months for each intern hosted. IMLS funding supports a stipend for each intern that is sufficient to cover living expenses and travel to and from the

host city. Over the past two years, students have worked on a wide variety of projects, such as:

- Supporting the NSF Blue Ribbon Task Force on Economically Sustainable Digital Preservation and Access at OCLC
- Assisting teachers using Archive-IT for the Internet Archive's K–12 Web Archiving Program
- Examining *Trusted Repositories Audit & Certification* (OCLC and Center for Research Libraries 2007) reports at the Center for Research Libraries
- Working with LOCKSS/CLOCKSS to increase publisher participation
- Performing audio file assessment, identification, and conversion at the Smithsonian Institution

To prepare for these summer internships, during the winter term each student enrolls in courses specific to the substance of the internship project, so they have basic knowledge about aspects of digital preservation that they can bring into the workplace.

The IMLS "Engaging Communities" grant also enabled us to create a special internship course, Practical Engagement Workshop in Digital Preservation, which provides scaffolding for digital curation interns working in the southeast Michigan area at sites with and without strong mentorship. The course is centered on the specific projects being undertaken by the students, providing instruction in these areas covered by the local internships and engaging the students in group problem solving. The classroom experience equips the students with some of the organizational skills that they require to do internships in a wide variety of cultural organizations to bring digital curation perspectives to non-profits that possess digital information but little expertise in managing it. Sample projects for the term interns include assisting MATRIX: Center for Humane Arts, Letters, and Social Sciences Online at Michigan State University cre-

ate and document digital curation policies; writing an RFP for digitizing materials at the Michigan Theater (a local cultural center and historic theater dating from 1928); assessing previously digitized materials at the Ypsilanti Historical Society; working with staff at the University of Michigan Libraries to draft a disaster recovery plan for HathiTrust Digital Library; and developing a digital preservation plan for faculty podcasts for the University of Michigan Archives, based at the Bentley Historical Library. Overall the experiences have been very positive, and students have gained experience and perspective on digital preservation issues. In the words of one student, "The content, it makes me think about digital preservation from the top down and from general to specific."

The design of the internships is based on a cognitive apprenticeship model with the goal of strengthening each student's ability to "think like a digital curator" (Yakel, Conway, & Krause, 2009). Overall the internships help students explore digital curation issues in practice. Student interns also get a chance to see how classroom theory and techniques are applied in the real world. In addition to students bringing course-based knowledge to the internship, they also bring internship-based knowledge and experiences back to the classroom, which enriches learning for all the students.

Conclusions

One of our goals for the well-established Archives and Records Management curriculum has been to deliver the major concepts associated with digital curation in an archival context and to help students work within a lifecycle framework to address preservation and access problems for both analog and digital information. Our review of the content of the ARM specialization led us to propose and then develop a separate Preservation of Information specialization. With this new educational program in place, we turned our attention

first to the delivery of that content in terms of pedagogical methods and instructional materials and then to the integration of technologies that support the curriculum's learning objectives to prepare students for professional practice. In this stepped process, we have kept in mind findings from the literature on digital natives, both in terms of their learning styles and their facility, or lack thereof, with technology. The results of this process are still in development, but our evaluation plan is targeted to developing evaluation instruments for examining students' adaptation competencies with learning new tools, archival-curation conceptual frameworks, and meaningful use of the tools to meet classroom learning objectives.

Acknowledgements

References

Bennett, S., Maton, K., & Kervin, L. (2008). The digital natives' debate: A critical review of the evidence. *BJET, 39*(5), 775–786.

Consultative Committee for Space Data Systems. (2002, January). *Reference Model for an Open Archival Information System (OAIS) (CCSDS 650.0-B-1, Blue Book).*

Lee, C. (2009). Matrix of digital curation knowledge and competencies (Overview), (Version 13). Retrieved from http://ils.unc.edu/digccurr/digccurr-matrix.html

Lee, C. A., Tibbo, H. R., & Schaefer, J. C. (2007). Defining what digital curators do and what they need to know: the DigCCurr project. In *Proceedings of the 2007 Conference on Digital Libraries* (pp. 49–50). New York, NY: ACM Press.

Oblinger, D. (2003, July/August). Boomers, gen-Xers and millennials: Understanding the new students. *Educause Review, 38*(4). Retrieved from http://www.educause.edu/ir/library/pdf/erm0342.pdf

OCLC and Center for Research Libraries. (2007). *Trusted repositories audit & certification: Criteria and checklist.* http://www.dcc.ac.uk/resources/tools-and-applications/trustworthy-repositories

Sharpe, R., Beetham, H., Benfield, G., DeCicco, E., & Lessner, E. (2009, May). Learners experiences of e-learning synthesis report: Explaining learner differences. Retrieved from https://mw.brookes.ac.uk/display/JISCle2f/Findings

Yakel, E., Conway, P., & Krause, M.G. (2009). Thinking like a digital curator: Creating internships in the cognitive apprenticeship model. In H. R. Tibbo (Ed.), *Digital curation: practice, promise & prospects: proceedings of DigCCurr 2009, April 1–3, 2009, Chapel Hill, NC, USA* (pp. 7–11). Chapel Hill, NC: School of Information and Library Science University of North Carolina at Chapel Hill.