Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice

PATRICIA GALLOWAY

RÉSUMÉ Depuis 2001, l'auteure donne un cours de deuxième cycle en archivage numérique à la School of Information de la University of Texas à Austin. Ce cours fait partie d'une suite de cours en archives numériques et en gestion de l'information qui s'insère dans les programmes en études archivistiques et en conservation. Ce texte trace les grandes lignes du développement de ce cours dans le but de décrire l'émergence d'une approche basée sur la pratique active et réflexive pour enseigner l'archivage numérique, et ce dans le contexte d'un environnement de laboratoire qui comprend un dépôt numérique fonctionnel. Trois aspects majeurs se dégagent de cette approche : 1) une méthode pédagogique qui appuie les étudiants dans leur apprentissage actif en leur permettant de compléter des projets archivistiques réels (la pratique réflexive dans le cadre d'une équipe à compétences variées); 2) une pratique qui vise à fournir une expérience en évaluation, en développement et en amélioration des méthodes de conservation dans le contexte d'un dépôt fonctionnel (des études de cas séquentiels); 3) le moyen de comprendre et de prendre en ligne de compte l'évolution même d'une infrastructure de conservation dans le temps (le modèle *Seed-Evolve-Reseed*). Enfin, les éléments primordiaux du contenu du cours comprennent l'histoire de l'informatique et de la recherche en matière de conservation, les méthodes et ressources pour la technologie de recherche, ainsi que les méthodes pour assurer la conservation à long-terme des objets numériques authentiques.

ABSTRACT Since 2001 the author has been teaching a graduate course in digital archiving at the School of Information, University of Texas at Austin, as part of a suite of courses on digital archives and recordkeeping, embedded in larger programs of archival studies and preservation studies. This essay outlines the development of this course over time in order to describe the emergence of an approach to digital archiving teaching through active and reflective digital archiving practice, in the context of a laboratory environment including a functional digital repository. There are three major aspects to this approach: 1) a pedagogical method to support students in active learning through carrying out real archiving projects (reflective practice in a mixed-skills team setting); 2) a practice aimed at providing experience in testing, developing, and improving preservation methods over time in the context of a working repository (sequential case studies); and 3) a means for understanding and accounting for the evolution of the preservation infrastructure itself over time (the seed/evolve/reseed model). Finally, major elements of course content include the history of both computing and

ARCHIVARIA 72 (Fall 2011): 169–196

preservation research, methods and resources for researching technology, and methods for assuring the long-term preservation of authentic digital objects.

Digital Archiving as Education and Research

It is a truism of archival education that in order to carry out their functions, all archivists are supposed to know their holdings, how and by whom they were created, valued, and used in their original and subsequent contexts, and how to protect them for the future. Protecting them for the future requires not only familiarity with the contents of collections but also with the technologies of the media that bear those contents, and the technologies by which the contents were inscribed on the media in question.[1] This assertion sounds straightforward when applied to familiar media and technologies of inscription like paper, pen, and ink, but as soon as we move to a somewhat less direct inscription technology such as the typewriter or especially the computer, we may ask whether the inscription technology itself is simply a better way of making readable marks. Conservators would point out that in order to remain accessible, all technologies of inscription (including not only the devices but also the substances or forces deployed in the inscription process) would probably need attention sooner or later. Historians would agree, since every technology of inscription implies a host of practices and even whole industries that may have a significant impact on the creation of the contents in the first place. It seems that archival education in general must be concerned with the medium of the archival record not only as providing structure, but for its role in framing context and even in contributing implicit content.[2]

As media become readable by the human eye only through the use of digital display technology, the issues of technological mediation become increasingly urgent. Where materials are digital, archivists must understand the technology of the environment of creation and the environment(s) of subsequent use, as well as the technologies used from that time onward to preserve and access them. In addition, they ought to preserve current and historical knowledge of these enabling preservation technologies. This necessity is even more complex than it sounds. Because digital inscriptions are generally made and reproduced using mass-produced equipment, the inconsistent standardization of that equipment

1 The phrase "technology of inscription" is elaborated by Katherine Hayles in *Writing Machines* (Cambridge, 2002), p. 24: "to count as an inscription technology, a device must initiate changes that can be read as marks."

2 The phrase "media materiality" has emerged in the discourse around new (read digital) media to refer to such a meaning/contributing aspect of media, not unlike the attention accorded to specific kinds of marks, choice of paper, and even styles of penmanship in diplomatic analysis. For the importance of digital media materiality, see Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, 2009).

has ironically meant not less variation, but more. To characterize the technology of inscription in such a way as to make its product "preservable," it is necessary to know information about the hardware, operating system software, and application software (that which seemingly does the "inscription") used by the creator, all of which may appear in different configurations, changing more or less rapidly and at different rates, even over the active life of the digital object of interest.

Every digital preservation task, therefore, will be as unique as the digital inscriptions being preserved. Although standards hold out some hope for easing the task for institutional archives – where the inscription environments may have been more or less tightly standardized (with periodic revisions) – this will not be the case for the inscriptions of individuals, which account for the majority of the materials held in collecting archives and will likely continue to do so. As a consequence, mastery of the full repertoire of inscription devices and environments that account for the creation of digital objects kept in archival repositories will always be an unfinished task; knowledge supporting the digital archiving task will always lag behind the need to do so; and learning digital archiving through doing will therefore always be compulsory (and will continue to be a feature of digital archiving practice). Accordingly, a process of learning that actually accomplishes work should be recognized not only as an acquisition of skill, but as the research that is often required to address unique cases. Education for digital archiving and preservation, therefore, has to be seen as educating for a whole set of uncertainties, beginning with the active life of the digital materials, and encompassing the series of unpredictable technological infrastructures that will sustain and display them.

Teaching Goals for Digital Archiving

This paper describes the ten-year development process of a course in advanced digital archiving theory and practice, formerly entitled "Problems in the Permanent Retention of Electronic Records" but soon to emerge as "Digital Archiving and Preservation." "Digital archiving" as used in this paper and in the course's new name is defined as "the practice of preserving (long-term or indefinitely) authentic digital cultural objects for present and future use."[3] The indetermi-

3 Patricia Galloway, "Digital Archiving," in *Encyclopedia of Library and Information Sciences,* 3rd ed., eds. Marcia Bates and Mary Niles Maack (New York, 2010), volume 1, pp. 1518–27, DOI: 10.1081/E-ELIS3-120044332. The 2008 report of a Library of Congress symposium on future directions in preservation (see Dianne van der Reyden, Diane Vogt-O'Connor, and Karen Motylewski, *Preservation Education in the 21st Century*, http://www.loc.gov/preserv/symposia/educrep.pdf, [accessed on 17 April 2011]), observed that, "Preservation in the digital context is variously called digital archiving, preservation, digital assets management, and digital curation …" (p. 8). The use of the term "digital archiving" in this article emphasizes the fact that the work described was developed at the nexus of

nacy of the digital archival environment does not mean that learning will be impossible or that general strategies and tactics cannot be taught. But it does imply that students of digital archiving need to understand past technologies of inscription, and the cultural context in which they emerged and were used, since it is unreasonable to assume that the majority of archival repositories can avoid responsibility for digital objects received on non-current media.[4] Thus it is especially important that students be introduced to the history of the practice of digital archiving: how the practice has developed, its successes and failures, and what has actually been done to preserve collections. This is necessary because the long-term preservation task will always rest on the activities of prior custodians.

Notable examples of changes in ideas about digital archiving as a result of the test of applied practice are not hard to find. In the 1990s, most people's access to a computer was still in the workplace rather than at home, and digital object formats were relatively simple. Research projects at that time suggested that it would be adequate to convert archival documents to a small number of standard formats that could preserve the restricted number of properties that were believed to have made those documents worthy of preservation, and then to migrate them on demand when needed by users.[5] By 2010, archivists were confronted by mass access to digital inscription technologies in the home, the challenges of complex digital objects and new media, and a consistent failure to reduce the preservation task by defining widely applicable sets of significant properties to enable the conversion of file formats to more "preservable" ones. Some form of emulation of the original supporting environment had begun to be recognized as not only desirable but a requisite part of the preservation of more complex, modern formats. The idea of simply converting everything to a single format was soon broadly rejected, given the recognition of the number of existing formats – even in restricted environments – and the impossibility of converting all of them.[6] A final example shows how quickly technological

preservation and archiving education, but from the more public perspective of the archives and the broader responsibility of archives for holdings that will be used over an unspecified period.

4 Given that personal archives are seldom deposited until a late phase of their creators' lives, at least some digital formats and media from these collections are likely always to trail current formats and media by as much as thirty years or more.

5 The Australian state of Victoria decided in the mid-1990s to use PDF and TIFF as the only preservation formats. See Howard S. Quenault, "VERS: Practical Digital Preservation," *Document numérique* 2/2004, vol. 8, pp. 23–35. During the same timeframe, the US National Archives and Records Administration (NARA) accepted only a very few file formats for preservation; the UK CEDARS project developed the concept of migration on demand in the face of arguments that periodic migrations of entire archives would be infeasible. See Maggie Jones, "The Cedars Project," *Library and Information Research*, vol. 26, no. 84 (2002), pp. 136–76.

6 The 2006–2009 InSPECT Project (Investigating the Significant Properties of Electronic

change alters the situation, as the developing technology of the World Wide Web has moved website structure from static to dynamic. In 2001 it made sense for the Internet Archive to harvest mostly static web page snapshots documenting the 9/11 attack on the United States, but by 2009 the ubiquity of dynamic websites containing streaming video and sound, consisting of objects managed in a subsidiary database and only assembled at the time of viewing, had led to broadly based international projects designed to achieve the ability to preserve all the features of dynamic websites in functioning form.[7]

Because of these complexities and persistent changes in technology, I believe that students need to master a core of the technology involved in digital archiving: a basic history of computing, both as a technology and as an infrastructure of inscription, and consequently of work and communication; current and historical hardware; operating systems and their file systems; and applications and their associated file formats. They also need to learn how to research unfamiliar technology, using both available information and testing methods. Finally, they need to learn how to assure the preservation of the digital object without change, from the time it comes into the possession of the archives. With these requirements in mind, therefore, it is necessary to provide students with knowledge of the international digital preservation community and its ongoing problem set. But it is also necessary to do more: to provide them with a

Content over Time), sponsored by the UK's Joint Information Systems Committee, concluded that while the concept of significant properties was fundamental to the workflow of digital curation and especially the generation of derivative copies in more recent formats relevant to specific user communities, and that while much further research on the implementation of significant properties definition is warranted, it was not realistic to assume that isolating a single timeless set of significant properties for any given format or medium would permanently be effective for digital materials kept over the very long term. See the final report at http://www.significantproperties.org.uk/inspect-finalreport.pdf (accessed on 7 July 2011). See also Jerome McDonough, Robert Olendorf, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Rachel Donahue, Andrew Phelps, Christopher Egert, Henry Lowood, and Susan Rojo, *Preserving Virtual Worlds Final Report* (31 August 2010), https://www.ideals.illinois.edu/handle/2142/17097 (accessed on 7 July 2011). Overlapping the 1990–2010 period, understanding of the documentation of work had evolved from the classic paper flows of hierarchically organized business described by JoAnne Yates in *Control Through Communication: The Rise of System in American Management* (Baltimore, 1983), to the collaborative digital networks described by Peter Botticelli in his classic paper "Records Appraisal in Network Organizations," *Archivaria* 49 (Spring 2000), pp. 161–91.

7 The September 11 Web Archive has been harvested by the Internet Archive as commissioned by the Library of Congress; it currently consists of more than five terabytes, delivered to the viewer through the Wayback Machine using automated Javascript recoding; see http://september11.archive.org/welcome.html (accessed on 7 July 2011). Currently the International Internet Preservation Consortium is supporting the development of the Web Archive (WARC) format, based on the Internet Archive's crawler output format, for conversion of websites. Two current projects, the Living Web Archives and the World Wide Web of Humanities, are being carried out in Europe to develop a capture format that retains the visual and interactive features of websites for digital discovery and archiving.

supportive environment in which to research, experiment with, and use digital archiving techniques in order to acquire the skills and understanding to work independently with digital objects, as well as the ability to carry out experimentation to recognize and solve the novel problems that they will inevitably encounter.[8]

How to do this? What should such an environment look like?[9] I have drawn on two models sequentially: first, computer programming team practice as developed in software engineering pedagogy, and second, the architectural design studio approach as adapted for engineering education in the 1980s and 1990s.[10] The crucial step is to create teams whose members have among them a group of skills suitable to tackling the particular project at hand, and who are supported by faculty and staff as they solve problems by expanding their knowledge and practice. The work of the teams is nevertheless constrained by a standard of outcome based on available infrastructure and preservation requirements. As will be discussed, we began modestly with internal projects from colleagues within our school, and then broadened our work to include projects brought to us by local archives; in all cases students have had a client external to the class itself, with a real archival goal to be achieved. In addition to the concerns of the client, the students have had to work within the constraints of the repository software we adopted to meet the challenges of capturing and stabilizing within the repository whatever digital objects were presented on whatever media and in whatever format. Finally, students have also had to organize themselves

8 An anonymous interviewee observed in 2001 that, "At a minimum, you need [professionals in] three [areas]. One is an archivist who knows what the requirements of the work are and the nature of the objects they're dealing with. One is a computer specialist who can mount and maintain applications. And the other is an … 'archival engineer' who has a combination of knowledge from the archives side and the IT side" (p. 221; quote is unattributed, acquired in a set of interviews of InterPARES participants by Michèle Cloonan and Shelby Sannett, "The Preservation of Digital Content," *portal: Libraries and the Academy*, vol. 5, no. 2 [April 2005], pp. 213–37). The argument in this essay is that the archivist can *become* an archival engineer and that a more extensive knowledge of technology than is implied in the interview is needed to complete the whole task of preservation.

9 When I began this work, there were no real models available. Even today, the DigCCurr Project and the subsequent Closing the Digital Curation Gap Project have not as yet (early 2011) arrived at a conclusive curriculum for digital curation: "curriculum just being developed..… Core content unclear; range of content is unclear.… Everything above is exploratory and experimental.…" See Helen Tibbo, "Educating the Curator: Digital Curation Education in the United States," presentation at the European DigCurV meeting, London, 17–18 January 2011. I would argue, however, that digital archiving education needs to *remain* exploratory and experimental, certainly at the graduate level.

10 The former is based on Frederick P. Brooks, *The Mythical Man-Month: Essays on Software Engineering* (Reading, MA, 1975); the latter is summarized in Brian E. Thompson, "Studio Pedagogy for Engineering Design," *International Journal of Engineering Education*, vol. 18, no. 1 (2002), pp. 39–49, where it becomes clear how similar the requirements of digital archiving are to those of an engineering discipline.

within a small set of roles to accomplish the work, given their skills, schedules, and dispositions. A decade of experience with this course has convinced me that hands-on work with real digital records and real digital archiving practices, where students are challenged not only to learn and use what is known but to push the boundaries of what is known, is the most effective way to create a setting in which this learning can happen.

Infrastructure Requirements

Given the above goals, it is necessary to build both a suitable archival repository within the educational setting to hold digital materials, and an ancillary security and protocol infrastructure for the handling and management of original materials under preservation care in order to protect them against potential damage during initial processing. The repository requires adequate IT support and laboratory space. Computer equipment to carry out pre-deposit processing is required to support students' participation in both experimentation and actual preservation work. Because it is desirable that the repository be taken seriously as an archives within its institution, it is especially important that this infrastructure become a vital part of the educational institution, serving other purposes in addition to the support of pedagogy and therefore entitled to draw on the institution's resources.

If digital archiving competence can best be learned through the archiving of actual digital objects, and if doing that constitutes research, it would be wasteful not to preserve adequate records of that research and practice together with lessons learned (for good or ill), as well as to pursue any unsolved or inadequately solved problems repeatedly until an acceptable solution is found. To that end, it is important to develop a practice of preserving and making available, not just the processed archival materials, but also the detailed reports on work done and protocols developed. This means that the repository must be designed to support and maintain both archived materials and the records of their archiving – this is no less than should be expected of any archival repository.

Theoretical Basis: Creating a World Within the Digital Preservation Arena

Building such an environment for education and research has incrementally incorporated the features of the real social arena in which digital archiving has so far been experimented with and has taken place.[11] In the digital archiving arena – which can be defined as those participants and groups committed to taking

11 I refer here to sociologist Anselm Strauss's concepts of arenas, social worlds, and negotiated order, as expressed by Adele Clarke in *Situational Analysis: Grounded Theory After the Postmodern Turn* (Thousand Oaks, CA, 2005).

an active part in the definition, practice, and institutionalization of digital archiving – multiple social worlds are represented. Classically, the most important stakeholders in the wider arena of cultural heritage maintenance are the creators, the custodians, and the users. In our case, the social worlds involved can be more specifically defined to include:

- Donors/creators of digital materials (individuals, groups, or companies);
- Digital objects of all kinds (the objects of all this attention);
- IT manufacturers producing hardware, software, and services;
- Archives, together with archivists and constituting one locus for producing the digital archiving literature;
- Users (scholars, artists, writers, and the general public);
- Archival educators, acting to carry out research on process (thus constituting another locus for producing the digital archiving literature) and to perpetuate effective practice;
- Archival studies students, learning about all the other participants in the arena.

In this context, the laboratory course we have created and the practices within it are a specific microcosm of the digital archiving arena, negotiated over time to incorporate:

- The endogenous theoretical discourse of the archival world about the content, context, and structure of archival collections;
- Digital archiving research outcomes as reported in an international literature;
- A hybrid institutional infrastructure, incorporating:
  - The School of Information as an education and research unit in the University of Texas;
  - A broad range of students interested in archival studies, preservation administration, digital libraries, and human-computer interaction;
  - The Open Archival Information System (OAIS) model for a digital repository, devised by an international group including government archivists and scientific researchers;
  - DSpace, an OAIS-compliant, open-source repository system devised initially by a partnership between a university (Massachusetts Institute of Technology [MIT]) and an information technology (IT) manufacturer (Hewlett-Packard), but now an open-source project supported by the Duraspace Foundation;
  - A host of demands on the practice of digital archiving generated by creators, archives, and users;
  - Funding providers (institutional, government, and private).
- Active local archives of various kinds, with digital archiving problems arising from their digital collections;
- The graduate university pedagogical environment of information

science and its literature, including research on computer-supported collaborative work and engineering education;
- Legacy practices from the computer science and electrical engineering worlds, as represented locally by the Austin Goodwill Computer Museum, its volunteers, and its own concerns for hardware preservation;
- The digital forensics community and its forensic capture and authentication practice.

The following section outlines the sequential development of this course as a historical process that tracked both developments in digital archiving research and the challenges of learning to function in a context of incessant change.

Evolution of a Model of Learning and Research

The learning and research model did not emerge fully formed; it began as the product of a lot of experience in archival practice and teaching. I brought to the task twenty years of experience at the Mississippi Department of Archives and History where, beginning in 1980, I created an IT infrastructure for the general activities of the institution. There I learned how to take advantage of archivists' intelligence and passion for their jobs to introduce computer technology gradually through demonstrated efficacy for particular tasks, and a participatory framework in which representatives of several divisions of the institution had a say in the process. Fifteen years later, as increasing amounts of born-digital government records were beginning to challenge archivists, the time was right, and staff members were confident enough for us to explore and establish the basis for a digital archival function at the Department.[12] In 2000, I joined the respected archival education program built by David B. Gracy II at the University of Texas at Austin, where I was hired to create a digital archiving specialization at the union of the existing archives and preservation programs. The advanced digital archiving course (INF 392K) is therefore embedded in a full archival studies program that meets the Society of American Archivists' Graduate Program in Archival Studies guidelines, including: a two-semester course sequence on the basics of archival theory and practice; an archival history course; an archival appraisal course; a records management course; and two other digital courses that I also created, one an introductory course that is

12 The research and development that lay behind the creation of a digital archives for Mississippi was funded by the National Historical Publications and Records Administration in 1997–1999; see the 2000 final report, *Mississippi Electronic Records Initiative: A Case Study in State Government Electronic Records*, http://www.ischool.utexas.edu/~galloway/nhprcfinrept.doc (accessed on 7 July 2011). Clearly it would have been much less likely for me to imagine the creation and operation of a digital archival repository in Texas had I not already done this in Mississippi.

limited to work on the students' own digital materials, and the other a detailed metadata seminar. As does the whole of the archival program, the course also benefits (and attracts students from) the preservation administration program also supported at the School of Information.[13]

Over the course of ten years, 143 students have taken the advanced digital archiving course (INF 392K); working in teams, they have completed over forty-one archiving projects.[14] Because of my interest in exposing students to real digital materials (many under threat of obsolescence) that needed to be archived and thereby secured,[15] the focus has been on the first steps in the archiving process, i.e., the capture into a secure repository of authentic digital objects that are adequately described. Projects have been drawn from the School of Information itself (creating a repository of material documenting the history of the school), and from campus and other local archival repositories for which our project has served as an incubator for their own early steps in digital archiving. Class size has ranged from four to twenty-four as the course has developed, and the number of projects per year has ranged from one to seven. Although several different courses were developed early on, and elements of the course described here were present in the first overview course I taught in 2000, the full-blown concept did not emerge instantly.[16] Instead, the model I now use has evolved through our actual work with archiving materials into, and managing, a departmental institutional repository. It was certainly inspired initially by the software engineering model expounded by Frederick P. Brooks, Jr. in *The Mythical Man-Month* and the activity theory-based critique of that model offered by reports on a project I participated in as an anthropology student with Dorothy Holland in 1993.[17] It

13 Although the UT-Austin School of Information's conservation program was terminated in 2010 for lack of funding, the preservation administration program continues to attract students and to partner well with archival studies.

14 The author would like to thank all the students who have taken the course, cumbersomely named "Problems in the Permanent Retention of Electronic Records," which they came to refer to as simply "Problems" (to be renamed in 2012 as "Digital Archiving and Preservation"). Their resourcefulness, creativity, and love for the work have been indispensable ingredients in the development of the course.

15 This emphasis also reflects the major foci of research in digital archiving generally, which so far has not moved much beyond trying to solve the problem of stabilizing digital objects in a preservation environment.

16 The initial course's syllabus can be found at http://www.ischool.utexas.edu/%7El389c5pg/ and links to the full list of versions of the "Advanced Seminar on Long-Term Preservation of Digital Objects" can be found at http://www.ischool.utexas.edu/%7Egalloway/syllabi.htm (both accessed on 7 July 2011).

17 See Brooks; Dorothy Holland and J.R. Reeves, "Creativity and Rationalizability: Beasts in the Tar Pits of Software Engineering," paper presented at the annual meeting of the Society for Literature and Science, Atlanta, GA, October 1992; Dorothy Holland and J.R. Reeves, "Activity Theory and the View from Somewhere: Team Perspectives on the Intellectual Work of Programming," in *Context and Consciousness: Activity Theory and Human-Computer Interaction*, ed. Bonnie Nardi (Cambridge, MA, 1997), pp. 257–82; Patricia Galloway,

was also very much affected by the evolution of ideas and standards in the digital archiving community of practice itself.

Initial Elements of a Program: OAIS Repository Model, DSpace, and Multi-Skilled Teams

During its initial two years (2001–2002) the digital archiving class studied the Open Archival Information System (OAIS) model, and developed a database and interface to attempt to mirror OAIS using the LAMP open-source software set (Linux, Apache, MySQL, and PHP) on a discarded Intel-based server. Students, initially consisting entirely of those with an archival studies specialization, were each asked to develop individual "domain expertise" to support team-structured work by a single team; our aim, however, was to integrate team work in the process and to include cross-training so as to ensure that students gained a broader range of knowledge about those particular skills than any of them possessed before they began the class.

MIT and Hewlett-Packard released the DSpace open-source repository software as a public project in the fall of 2002 and we adopted it as a platform in the spring of 2003. We evaluated the DSpace software's capabilities in the context of emergent requirements for permanent digital archiving and gained experience using DSpace for our own archiving of class work. The problem of email management in government where multiple systems were in use, and the problem of permanent preservation of the final digital typesetting files of academic journals, laid the foundation for using the DSpace repository as a model to support applied research in digital archiving for institutions outside our own. To the division of labour into domain expert areas we added a practice of project meetings in which each expert offered a weekly summary of work. During these years when classes were very small, it also became clear that classwork-as-discovery was an effective way for students to work.

From the beginning I was committed to creating a set of courses that I would teach for at least fifteen years; my experience in the Mississippi archives had demonstrated that a significant effect could not be achieved in a short time. Having already tracked the digital archiving and preservation literature closely for nearly thirteen years, I was convinced that the development of the field would continue at an uneven rate.[18] It would thus be important to recognize that, as the

"Playpens for Mind Children: Continuities in the Practice of Programming," *Information and Culture* (2012), in press.

18 I had begun tracking the literature after having represented the Mississippi archives at the landmark 1991 meeting sponsored by the NHPRC as a project of the Minnesota Historical Society, reported in the publication *Research Issues in Electronic Records: Report of the Working Meeting* (St. Paul, 1991), and under the 1997–1999 NHPRC grant to the Mississippi Department of Archives and History to design and create a digital archives for the state,

course continued, we would be addressing multiple genres of content, different social and technological contexts, and multiple file and media formats, all in an environment of an emerging digital archiving practice. If, under these conditions, a course repeatedly revisited these issues, we would be able to develop an evolving group of student researchers with expertise that we could bank in the repository itself in the form of both archived materials and reports on the archiving process, which could in turn serve as a resource for both past and future students.[19]

In addition, the problems we had addressed in these first years already pointed to several topics of interest that bridged the familiar archival categories of content, context, and structure. Although we were thus far essentially self-archiving our own output in current digital formats, we realized from trying to think about other people's problems that it would always be necessary to deal with legacy digital creation environments, media, and formats. Especially in the realm of personal collections, intact fonds (consisting at a maximum of hardware, software, storage media, and user-created files) would be evidence of user adaptations both *to* and *of* their computing environments, and their own practice in the production of digital objects. The very schematic, non-prescriptive nature of the emergent OAIS model and the ongoing (and publicly available and observable) evolution of DSpace, pointed to the fact that the infrastructural system environment that would be so crucial to the very act of preservation would itself be changing constantly, since as a digital object the repository was subject to the same considerations that we were addressing for the repository contents.[20] We would, in short, have to accept and deal with the inevitable development of all phases of the digital archiving problem set over time. Finally, the literature in digital archiving (as well as that on the history of libraries and archives generally) made it clear that both descriptive and preservation practices would continue to evolve over time: there would be no permanent solutions.

which I directed. At the time of these early efforts in Texas, I was also serving on an advisory board for the NARA/NHPRC project "Methodologies for Preservation and Access of Software-Dependent Electronic Records" being carried out by the University of California–San Diego Supercomputing Center.

19 In a sense this already resembles the multi-year focus of design–studio practice that will be discussed later, in which beginning students start by working with simple tasks and move on to more sophisticated ones, while mentoring less advanced students who follow them into the same project (see Thompson). In our case, the support of more advanced students is enhanced by the investigations and discovery reports of prior students.

20 It should be pointed out that commercial repository systems are always evolving too, but it is not possible for the archivist to know exactly what is going on in the proprietary case.

Establishing a Departmental Institutional Repository and Incubator for External Projects

It had been clear from the beginning that growing the DSpace project space to become a digital archival repository was necessary for archival students to engage with the weighty responsibility of digital archiving in a real repository and working on real holdings. Our well-established program in archival studies had made colleagues across the School aware of the concerns of archival science and the advantages of archival management of their records. Thus in 2005, with a class size of nineteen, we made the commitment to construct and support an institutional repository for the School of Information, to include faculty records, the School's website in its historical incarnations, and a set of tutorials developed for the benefit of students by student lab assistants.[21] We began by archiving digital files, including publications, syllabi, web pages, and presentation slide sets, chosen by four faculty members to represent what they considered important materials representative of their careers. We also experimented with crawling and archiving the current School website. In addition, having been involved in an advisory capacity with policy decisions taken to acquire the collection of hypertext novelist Michael Joyce by the Harry Ransom Center (HRC), it was possible for me to engage the class in work on an initial group of materials from that collection, thereby providing a proof of concept for the archiving of this first, substantially digital collection for the HRC and beginning to develop a protocol for processing legacy formats. During this iteration of the course, students began to keep task journals to record the details of their developing digital archiving investigations and practice.

We also combined standard archival practice, familiar to most students in the class, with OAIS procedures as instantiated in DSpace to establish a basic workflow for handling digital archival materials.[22] The work thus begins outside the repository with an *inventory* of media and their contents (making use of media labels as well as directory listings), normally recorded in a spreadsheet accounting for digital objects at the individual file level. *Capture* entails write-protecting media, copying files into a secure holding space for transfer, and digitally recording file structures for inclusion in a documentation file.

21 Two of the students in the 2005 class would become digital archivists for campus archives: Catherine Stollar for the Harry Ransom Center (HRC) and Zachary Vowell for the Center for American History. The fact that these opportunities developed was contingent on the acquisition of digital collections, but once the two archivists were hired, inventories of existing collections by students undertaking preservation administration projects led to discovery that both institutions had already accessioned digital materials as part of non-digital collections, thus suggesting new projects that the class could take up in its incubator mode.

22 In general when it is discovered that an item is inappropriate for the repository for some reason, we carry out only limited appraisal in co-operation with the creator or custodian of the materials.

Arrangement encompasses the construction in DSpace of hierarchically arranged "containers," called community, sub-community, and collection, so as to model archival hierarchies; this step includes providing aggregate descriptive elements like scope/content and biographical/historical narratives.[23] *Accession* into the created structures in DSpace follows the built-in procedures included under "ingest," i.e., the transfer of files and metadata into the structures defined in the arrangement step; the metadata, normally collected during the inventory step, provides granular description using Qualified Dublin Core metadata elements. Finally, along with the content, students archive a *report* on the details of their processing steps, together with any extra materials developed during the processing phase. This sequence remains relatively intact to this day, although further activities have been added to several steps, such as batch ingest for large collections; bitwise imaging of source media followed by extraction of cloned individual files as standard practice to replace copying; and the creation of container structures that mirror media containers rather than logical categories, providing the basis for building multiple, virtual, logical collection structures.

From our work with both our faculty and the collections of creative writers from the HRC, we borrowed the collecting archives focus on the digital object creator as individual. Following the pattern of Brooks's software engineering course, we adopted a model of working with the content creator or archivist as client.[24] As part of the development of our in-house digital repository, we thus worked with colleagues from our school on collections that they had created. This is still an unusual approach for traditional archivists, accustomed to both receiving paper records to archive long after the creator has lost interest in them (indeed frequently long after the creator has stopped breathing) and following standard processing procedures under which they take upon themselves the power to shape collections. Although a participatory model was instantiated in the DSpace project at MIT (to the extent that DSpace was created to support community management of collections and self-archiving to a central repository), our project had not yet adopted such a model except to the degree that students self-archived their documentation work.

It was important to the authenticity of the student experience that the DSpace repository be securely established as an archives for the School, but it was necessary to find a way of thinking about it that would recognize the necessity for change over time. Taking into account the overtly evolutionary status of DSpace

23 At this point in the development of the course, arrangement followed paper archival practice in that it did not recognize media dependencies.

24 For a report on this project, see the case study from the 2006 New Skills for a Digital Era conference, Catherine Stollar [Peters] and Thomas Kiehne, "Guarding the Guards: Archiving the Electronic Records of Hypertext Author Michael Joyce," *New Skills for a Digital Era*, ed. Richard Pearce-Moses (Chicago, 2006), available at http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf (accessed on 7 July 2011).

as an open source project, together with our emerging sense that the contents of the repository would also be evolving through our care of them over time, we were inspired by the so-called "seed/evolve/reseed" model, developed by Gerhard Fischer and his colleagues at the University of Colorado Center for LifeLong Learning and Design, to underpin the development and ongoing management of the School repository as a collaboration among faculty, students, IT and other School staff, and potential depositors and users.[25] This model recognizes that "systems that evolve over a sustained time span must continually alternate between periods of activity and unplanned evolution, and periods of deliberate (re)structuring and enhancement" – in other words, it recognizes that social groups and the systems that support their collaboration must evolve over time, and robust social structures and systems must be able to evolve or they will be discarded.[26] From my point of view, the presence of a repository was vital: we needed to control it in order to have the freedom to experiment, and we needed to operate it over time in order to confront the problems of long-term digital preservation that would emerge. But it would also require investment in the form of IT support, and ideally that support would become a part of the School's infrastructure. Hence to be financially sustainable as well as functionally legitimized, it would have to have the participation and support of a good proportion of the members of the supporting organization who might make use of it for archiving, study, or teaching.

By the time we began work in the spring of 2005 on an archival institutional repository we thus had the *seed* in the form of the DSpace repository and the 2003–2004 student work toward intellectual property and preservation policies for its use. In spring of 2005 we had gained and recorded some experience with applying DSpace to the problems of digital archiving of multiple formats of digital objects. This work served as the first steps toward *evolving* the repository by adapting it to our use within the constraints of the software as we implemented it, while at the same time documenting our experiences within the repository

25 The model in fact recommends that initial design products be somewhat under-developed and left significantly flexible so that modification by a wide range of the participating community can be accommodated as the system goes into use. This has been characteristic of DSpace, whose original release was aimed at meeting needs in a specific environment under the OAIS model, while leaving some aspects of the model out of the original release, making others user-modifiable through the graphical interface, and as an open system making it possible for programmers to add any needed features, which can be submitted to the community of users for a decision as to whether they might be added to the system for all to use or remain as local usages.

26 See Gerhard Fischer and Jonathan Ostwald, "Seeding, Evolutionary Growth, and Reseeding: Enriching Participatory Design with Informed Participation," in *Proceedings of the Participatory Design Conference (PDC'02)*, eds. T. Binder, J. Gregory, and I. Wagner (Malmö, Sweden, June 2002), pp. 135–43, available at http://l3d.cs.colorado.edu/~gerhard/papers/pdc2002-ser.pdf (accessed on 7 July 2011). This paper in fact addresses using the SER model for a participatory collaboration system to support course instruction.

itself. Finally, the spring 2005 work and our observation of the use of the repository by other depositors and users over the summer and fall semesters were incorporated into a model for institutional use and archival preservation over the long term in order to *reseed* the ongoing repository project with synthesized findings for another round of student projects and research. This we undertook in spring 2006 after working with IT staff to update the DSpace platform, which had itself been evolving and reseeding on the basis of reports on work like ours. We then configured it according to our findings and migrated the contents to the new platform. Our actual work with the repository was even changing our views of the theoretical aspect of digital archiving as we grappled repeatedly with gathering materials, placing them in the repository, designing and adding metadata, and attempting to organize the repository for accessibility.[27]

Grant-Funded Research into Institutionalization: Studying the Problem of Sustainability

The work in 2005 taught us much about what the OAIS documentation refers to as the "Producer–Archive Interface." With the assistance of a grant received from Ingenta through the ALA Library Research Round Table, we began in 2005–2006, as part of the reseeding process, to examine our work carried out thus far by investigating how the project had been received by the faculty and staff participants, its impact on awareness of their work, and the tangible costs of the project – all to assist in deciding whether it would be possible to make the repository sustainable as part of the infrastructure of the school. The conclusions would be used to aid the "evolve" step.[28]

The measurable benefits studied were citation patterns and usage outcomes. Although faculty materials had not been available long enough to have been cited yet, we did find that deposits were beginning to appear in search engines. From 2001–2005, class size in the digital archiving course had grown from four to nineteen; students using the repository for their own projects had published eight papers and completed two capstone projects, four independent studies, and

27 See the Joyce project case study cited in footnote 24; also see Maria Esteva, "Text and Bitstreams: Appraisal and Preservation of a Natural Electronic Archive," another case study from *New Skills for a Digital Era*, available at http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf (accessed on 7 July 2011).

28 The author thanks Ingenta (now Publishing Technology) and the ALA Research Round Table for this grant, which made possible the pause and reflection needed to accomplish the next step in the evolution of the teaching environment. A brief report on this project, in which students Maria Esteva and Larry Stewart assisted in the investigation of citation patterns and cost-benefit analysis respectively, was included as a case study by Patricia Galloway entitled "The Eyes of Texas: What can Archivists Learn from Working with an Institutional Digital Repository?," a *New Skills for a Digital Era* case study available at http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf (accessed on 7 July 2011).

research for a dissertation. Measurable benefits therefore included curriculum enhancement, faculty and student scholarship, preservation of digital objects in the repository, and Web publishing of project findings. During the same period, repository costs amounted to seven thousand dollars per year, consisting of 5 percent of the system administrator's salary per year (six thousand dollars) and an incremental hardware cost per year of one thousand dollars; there were no software costs.

What we called "perceived benefits" of the study included the preservation value of the repository and the experience of those whose materials were archived. The repository provided for the archiving of scheduled records for the School, which state records law required the creating unit to preserve. It promised increased citations of faculty work. It preserved archived learning objects created by faculty members, which could be reused as such or used retrospectively to study curricular developments. It preserved our locally created software tutorials, which promised to be useful in the future to the task of software preservation and the use of legacy software for the preservation of older digital objects. It provided a platform for student projects in digital archiving. In addition, archived materials could provide for the study of the history of technology use in the School, for a longitudinal study of the personal information management of faculty members, and (through the archiving of documentation on archiving) for the study of the history and evolution of digital preservation practices. Finally, those faculty members whose materials had been archived welcomed another opportunity to work with students, an additional chance to share scholarly and pedagogical work, and the occasion to shape their legacy and to reflect on their past work.

The fourteen students who took the course in the spring of 2006 broadened the range of projects carried out for the School itself. They continued work on the School's website, experimenting with a crawler protocol that would concentrate on capturing only the postings that were required to be kept under the School's retention schedule; they archived more of the online software and hardware tutorials created by the School's IT lab assistants as well as multimedia documents that were part of a publication series sponsored by the conservation program; and they experimented with the treatment of email attachments from a body of email associated with a professional association. This work established as standard practice the use of student task journals as a formal deliverable of the course, further encouraging students to both record their practice and reflect upon it. It also demonstrated the value of revisiting formats and taking on more challenging archiving projects.

Sequential Case Studies through Genre and Format Repetitions

As we continued digital archiving work after 2006, we were able to undertake new projects consciously chosen to repeat genres or to continue work from previous years. Students were able to refer to previous work and take up the recommendations and problems described by former students while keeping up with the literature in the field. In addition, complex digital objects represented an important new problem set. Literary collections challenged us to cope with idiosyncratic file-naming conventions (an issue observed earlier, but for the first time much more dominant), variable access requirements to protect privacy, and corrupt and unreadable files. For the first time, we worked with the contents of an entire laptop and were challenged by the fact that it was recorded in an outdated backup format that presented great difficulty in discovering the details of the software used, even though we were able to locate the relevant hardware device. Ongoing faculty deposits required attention to copyright and journal embargoes; work with another faculty member who was preparing to retire provided an opportunity to plan several years' archiving work with him. Work on the School's tutorials took the step of intervening in digital record creation by providing recommendations on incorporating preservation concerns in the workflow of future tutorial production. There were also new media, formats, and partnerships. For the first time we archived commencement videos. Significantly, we carried out incubator work for the Briscoe Center for American History (BCAH), archiving two groups of digital sound recordings from a collection in the Center's new Videogame Archives.

Advanced Engagement with Legacy Media: Partnership with the GCM

The year 2009 saw twenty-three students in the class, which allowed us to undertake six projects. We did one project for HRC, which required the team to develop a protocol for access restrictions to confidential files. The other five projects were done for BCAH; one of them (the digital collection of a distinguished, deceased faculty member from our school) was quite complex, evidence that he had been just as innovative in adopting and adapting new digital technologies as he had been in his conservation research. The remaining materials from BCAH were again from the Videogame Archives, although we were able to repeat genres: music files made by the donor we had already worked with, where multiple versions of the same song were preserved; and email from a game designer, which introduced issues beyond the email itself because he had made use of all of the many other office functions of his email client. Another team took up a group of design documents created by the same donor, while yet another worked with a new genre – game computer code from the collection of a game producer – to which they were able to add information from interviews with her and with the lead programmer on the project.

The age of many of these materials, dating to the 1980s, forced us to look for hardware to make the initial capture of bitstreams, and in that process we began a fruitful partnership with the Goodwill Computer Museum (GCM) of Austin, a subsidiary of Goodwill Industries of Central Texas and beneficiary of its huge electronics recycling stream. The GCM and its volunteer engineers were able to assist us with assembling hardware and software to perform file capture, and this partnership enabled us to begin extracting disk images systematically for preservation, in this first case from 5.25" floppy disks. Students involved in these projects recommended that the ongoing repository project assist the GCM in preserving the knowledge of retro-computing specialists as well as available documentation for legacy software and operating systems.[29] This partnership developed further in the 2009 fall semester as a three-way partnership with BCAH, as digital archivist Zach Vowell worked with the GCM to develop a prototype, multi-platform server (nicknamed Frankenstein I) that we would put into hard use in the 2010 class.

Digital Archaeology Laboratory and the Establishment of Forensic Practice

I was able to secure the use of a research laboratory room for our growing collection of hardware, media, documentation, and other materials in the spring of 2010. The digital archiving class had twenty-four students and we did seven projects, continuing to explore examples of previous problems and tackling new ones. Three projects were carried out for BCAH, with all three being archived not on our server but in a new DSpace instance operated by the University of Texas General Libraries as the UT Digital Repository; this marked our first example of the transition of an incubator project to independent hosting, although we retained copies of student project documentation in our repository. We worked once more with materials from the game music creator, including email and what is now a relatively exotic format for audio files using the ADAT machine he had used to record them in the first place.[30] Students working on the ADAT project also turned to the donor as expert, and recorded and archived an interview with him about his work and use of the machine in practice. This work bore fruit when one of these students carried out a capstone project at

29 This recommendation has grown into two projects, one to assist the GCM with the development of an archive for support materials for their hardware collections, and another to develop methods and protocols for recording work practice in restoration and operation of legacy hardware. See Patricia Galloway, "Retrocomputing, Archival Research, and Digital Heritage Preservation: A Computer Museum and iSchool Collaboration," *Library Trends*, vol. 59, no. 4 (2011), pp. 623–36.

30 The Alesis Corporation's Digital Audio Tape recorder in question recorded eight tracks of digital audio onto Super VHS tape; see the article at http://en.wikipedia.org/wiki/ADAT (accessed on 7 July 2011).

BCAH to transfer the practices developed in the project, and to standardize them for both BCAH's in-house practice and our own future projects. We also captured from legacy media and archived for BCAH the transcriptions of the 1988 Presidential campaign oral history interviews using the old Nota Bene word-processing format. For a new incubator partner, the Alexander Architectural Archives, we captured and archived copious text materials pertaining to Mayan archaeology. We continued and nearly completed the archiving of existing School tutorials, added to an existing faculty archive, and worked with yet another retiring faculty member to initiate his archive. During 2010 we also met a significant goal of capturing almost all materials through disk imaging, and received funding under the University of Texas Temple Teaching Fellowship to purchase a forensic workstation to undertake the next stage of investigation of the usefulness of this methodology to the purposes of digital archiving.[31]

Case Studies and Reflective Practice

Pedagogical Value of Case Studies

I have come to view our project work as a series of case studies exploring the design space of digital archiving practice: testing recommended procedures and confronting them with the actuality of the process of archiving real collections. Because we have been working with the holdings of real archives, and especially as we have had the opportunity to work with currently-received collections, we have constructed our complex social world of digital archiving by bringing in archival tradition and research, involving ourselves with actual archival practice and its real-world demands through multiple institutional partnerships, and becoming acquainted with the practices of computer engineers and digital forensics experts to add to our archival work. Our practice has thus begun to approximate the kind of real and complicated experience that archives are facing with digital collections. Clearly we see this in our campus partnerships with collecting archives: because student teams have active archivists as clients, they are truly actors in that space.[32] As the work has proceeded we have increasingly

31 Catherine Stollar Peters and Gabriela Redwine, another former student who had followed Peters as digital archivist at HRC, were active in the international research that led to the report *Digital Forensics and Born-Digital Content in Cultural Heritage Collections* (CLIR report 149, January 2010, by Matthew Kirschenbaum, Gabriela Redwine, and Richard Ovenden); April Norris, a current PhD student who took the 2005 class, is doing research on this issue.

32 Although so far we have mostly worked with non-institutional materials, I have continued to maintain an interest in the Mississippi Department of Archives and History digital archives and the problems it is confronting. In 2010 the Texas State Library and Archives undertook a new initiative to build its own digital archives, which I hope will give us opportunities to work with the specific problems of governmental archives going forward.

pushed our efforts back in time into legacy environments and have become more convinced of the absolute necessity for permanent preservation of the original bitstream, from whatever era, where possible. Case studies have served two purposes pedagogically: as iterative investigations of a single generic problem or as extended exploration of a single authorial fonds. All case studies have brought up challenges for capture, arrangement, and preservation.

Using a series of case studies applied to the same file category[33] allows investigations to progress in understanding the category and its manifestations in time and space through repeated tasks. We employ a constant comparison approach, bringing in ongoing research in the field that applies to a particular body of materials, our relevant previous work, and our active experience as working digital archivists when weekly project meetings reveal problems and discoveries across projects. In this way, we have incrementally tried to pursue work on the categories of text, email, web pages, computer code, and sound, learning from our work and applying that learning to repeated work in the same area. Through that experience we have found that changes in software and software-produced formats under the categories we have addressed are far from linear, reflecting both the competitive motivations of commercial software vendors and the local modifications or preference settings of actual users.

Although many of our projects have been focused on specific file categories, especially relevant to systematic recordkeeping in government or business contexts that often reflect business functions, most have so far mirrored the more heterogeneous and idiosyncratic nature of personal recordkeeping by faculty members and by authors whose works are acquired by collecting archives. As such it is appropriate to think about each of them as a single fonds, or natural aggregate from a single source, and in each case we have treated them as such. In the case of our own faculty members, we have attempted to conceive of digital holdings as a distinctive output of the person, yet with awareness that it is only part of what s/he has produced. In the case of individuals whose digital materials are part of an already archived larger fonds containing paper materials as well, we have been able to work in a more focused way with an awareness of how the digital part of the collection fits into the whole.

Thus, projects were chosen with the idea that each would constitute a coherent set of materials in some sense, if only as "all the digital materials created by the donor" or "all the digital materials from a specific accretion." In this way, we considered each deposit in each case as an element of an individual's fonds and thus have been able to explore how this traditional archival concept plays out in the digital context. Further, as has been understood since archives first considered the acquisition of digital materials from individuals, the preferred way to obtain these materials is periodically and/or as soon as possible

33 I am using the term "file category" here to refer to functional type.

after they have gone out of use, so that the entry of digital materials into an individual's fonds in an archival repository should most frequently be by accretion during the lifetime of the donor.[34] Thus, as materials are iteratively received from an individual donor, students working on a later accrual to a collection are able to follow the donor's preferred structuring of groupings of files and its reflection of classification as ongoing process as well as the preferred file-naming convention. The degree to which the externalities of hardware and software adoption affect the progressively accruing collection introduces new challenges to archiving the materials. This approach has been particularly effective for us in the few cases of faculty recordkeeping where records donation has been iterative; it has also been reflected in the well-curated and well-known sequential acquisitions we have worked with from the HRC.

Fonds related to an individual, when captured in the holistic environment of the entire disk drive of the donor, raise many other questions as they allow for the capture and description of additional layers of creator intentionality if the retention of forensic images is allowed. We have as yet not done this kind of analysis using these tools, but the acquisition of the forensic workstation in 2010 has made it possible to undertake such a trial in spring 2011.

Finally, many of our case studies have challenged students with significant problems of media, file systems, and file formats for the task of capture. Obviously the physical medium needs to be read, so a peripheral device capable of accessing that medium needs to be available, and students have participated in tracking down and devising testing methods for suitable legacy media platforms. All physical media have some kind of file system imposed on them for the arrangement of the files, so for the medium to be read after copying or cloning also requires the availability in some form of the system (operating system or other) used to format and/or manage the medium.[35] Although it is not necessary for capture that original software be present to open, read, and display the contents of files (and indeed it might compromise authenticity to use it), such software (or more limited software readers or emulation thereof) needs to be available in some cases to provide access to the files. To date most students have

34 Lucie Paquet argued for this practice as early as 2000 in "Appraisal, Acquisition, and Control of Personal Electronic Records: From Myth to Reality," *Archives and Manuscripts* (November 2000), pp. 71–91. For more recent advice to individual recordkeepers, see the booklets from the Paradigm (http://www.paradigm.ac.uk/workbook/index.html) and InterPARES (http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf) projects (both accessed on 7 July 2011).

35 It is not necessary to have a system that recognizes either the file system or the files themselves in order to capture digital files. Files can be discovered as chunks of information even if the file system is unknown by capturing the magnetic flux patterns recorded on the medium and analyzing them. Our projects will provide copies of legacy formats for testing to the Goodwill Computer Museum, which has an ongoing project to construct equipment to perform this task.

engaged with one of these related problems and all have heard them discussed in the concrete terms of actual projects in class discussions. They have also used file format registries and metadata extraction tools to identify file formats and acquire metadata for insertion into the repository's catalogue.

As the concerns of digital archivists have shifted from conversion and standardization to bitstream preservation and emulation, none of these requirements for capture have changed; the demands of preservation, however, have become more stringent. At least from the viewpoint of leading collecting archives, it is no longer sufficient to capture content and context, and relax requirements on structure: at capture there is more interest that "everything be recovered," hence the interest in forensic methods and our move to incorporate them into our practice.

Reflective Practice

The notion of reflective practice as experimentation in problem solving, which has been foundational to the teaching of the class, was borrowed from the work of Donald A. Schön. Schön has demonstrated how reflection-in-action is characteristic of work in a broad range of fields where professionalism amounts to the ability to deal with unique and unpredictable events – this could be a canonical description of the daily experience of archivists.[36] All students come to the digital archiving course with some useful skills: by 2010, the majority of students in our school were equipped with better-than-average (if not expert) skills as computer users, and this holds true for the archives, preservation, and digital library students who most frequently choose to take this course. Many students (usually more than half) are registered in an archives concentration, and non-archives students have usually had at least a foundation course in archival studies. Students are encouraged to take the digital archiving course toward the end of their program, when they will have acquired a high level of archival skills along with an understanding of the theory that articulates their use. Non-archives students likewise bring additional skills and talents to the work, especially digital library students interested in new media, and preservation students who are involved in preservation reformatting, both of whom may be expert when it comes to par-

36 Donald A. Schön, *The Reflective Practitioner: How Professionals Think in Action* (New York, 1983). See also Schön, *Educating the Reflective Practitioner* (San Francisco, 1987). Schön's work has been repeatedly cited in the literature on archival practice and education because of its appreciation of the indeterminacy of the archival task and the importance of student engagement with real-world problem solving through reflection. Some examples include Caroline Williams, "Studying Reality: The Application of Theory in an Aspect of UK Practice," *Archivaria* 62 (Fall 2006), pp. 77–101; and Vilnius University Faculty of Communication, *Outline of Training Principles and Objectives* (Digital Preservation Europe, 2007), http://www.digitalpreservationeurope.eu/publications/reports/DPE_Outline_training_Deliverable_2_1.pdf (accessed on 7 July 2011).

ticular formats. Archives students who bring traditional historical skills to the class find themselves significantly engaged in tracking down the technological history behind legacy media and formats.

As mentioned, the policy from the beginning has been to help all students benefit from one another's talents by creating multi-skilled teams, and all students are expected to make the best use of what they know in their individual projects. A first task is for students to inventory their team's proficiencies and tentatively assign roles according to knowledge and skill. Throughout the semester, students are encouraged to share their skills and assignments, with the most proficient serving as leader for a given task and instructing the others. Keeping a journal of actions and thoughts in order to externalize reflection is often advised as part of education for reflective practice. As the course has developed, journaling has proved especially helpful in assisting students in raising discussions of how supposedly established practices may or may not be suitable for the problem at hand. It also enables the shaping of digital archives practice as teamwork by providing students with a way of reflecting on their joint practice and how it is or is not working. Examination of the journals as submitted by students at mid-term allows the instructor to see how students have constructed their teams and the roles of their members through interaction around the problem at hand so as to allow application of the members' unique skill sets. It also allows the instructor to intervene where necessary through commenting on individual journals and meetings with separate groups to mentor more appropriate problem-solving approaches and team interaction. This element of the course is still in development, but already has proved its worth to students in helping them solve problems.

Laboratory Issues

Having created a laboratory and formed an alliance with the Goodwill Computer Museum, we have begun to craft a collecting strategy for legacy hardware, software, and media. The strategy is based on a combination of our experience thus far, the inventories of digital materials in the repositories we have been working with, and discussions with the School's IT staff. Adding the forensic workstation for current computing environments will enable us to cover a relatively long temporal range of computing environments; we anticipate upgrading the current workstation with newer software and, eventually, hardware in the future. We expect to fill in gaps in legacy materials by further developing the Frankenstein project by working with the GCM to build a hybrid machine that includes the capacity to interface multiple peripheral devices for capturing digital materials from legacy media; one of the class teams in the spring 2011 class was tasked with formalizing laboratory protocols for file capture and preparing a requirements document for a proposed Frankenstein II project. As a result, we are also considering a second "viewer" machine equipped with an

array of actual legacy processors together with virtual workspaces so as to accommodate multiple operating systems.

In the process of carrying out our reflective practice in digital archiving classwork, we have come to realize that all digital environments are highly complex; even well engineered hardware systems are bound to be unpredictable when operating systems and software applications are stacked on them. Further complications arise because these components are constructed by different and competing businesses. The challenges include: continuous changes in media, usually providing higher capacity in smaller space; changes in operating environments as system vendors seek to add new features and increasingly hide the mechanics of operation from the user; unpredictable user customization, provided for, but not dictated by, operating system vendors and encouraged by application software vendors; and specific user practice in the use of hardware and software, in each case as individual as the users themselves. The upshot of all these uncontrollable variables, as suggested at the outset, is that every collection of digital objects will be unique, and every digital archiving project will include research directed at media and preservation matters. We have found, as archivists always have, that to the extent that we document our own reflective work, we can acknowledge the work of those who preceded us in handling and documenting similar, though not identical, materials.

Lessons Learned and Problems Unsolved

Some of the most important lessons learned have been through our repeated work with the personally archived materials of individuals – both academics and artists. Initial issues have to do with the relationship between archivists and donors of digital materials; increasingly, donors are still alive at the time of deposit. This is especially pertinent to the progress being made toward the perfect capture of the intended final version of a file through the use of forensic techniques, since these techniques also enable the examination of unintended remains of erased files. It might seem obvious that the donor means to bestow only those materials that have been manifest to her at the time of deposit, using the digital tools available. Yet we are now facing new possibilities that have not yet been considered by donors, with the potential to alter the relationship between donors, archivists, and researchers. It has been routine for us to establish a "Submission Information Package Agreement" with individuals with whom we have worked, stipulating the conditions of deposit for both donors and archival repositories. As we have progressed to forensic capture, we must consider modifying the agreement to inform the donor of these new processing procedures and the possibilities for analysis they offer. The agreement should also specify the donor's wishes regarding those possibilities and especially for the application of preservation procedures to the deposited digital objects. Donors may allow archivists to preserve all files, even erased ones, on a given

medium, but may require that erased files be kept confidential longer than overt files, or released only under special circumstances that the archivist may have to arbitrate; donors may wish to have erased files expunged permanently, but only after they have reviewed them to see if they can recover files once thought lost. We expect to pay close attention to these issues and to consider what new conventions we may have to develop in order to clarify the archivist's role in this new relationship between donors, archivists, and researchers.

And how should we represent digital collections? We have seen an array of idiosyncratic arrangements intentionally created by individuals. The conventional archival tendency has been to make these arrangements more usable by researchers, often by obscuring the messiness of the received arrangement. In the digital environment, it is possible to have multiple arrangements, and we have begun experimenting with these possibilities but have by no means solved the associated problems. This issue will also call for more interaction with creators and a better understanding of their relationships with digital technologies of inscription. As a result of attending to directory structures, file naming conventions, and file creation dates, we know far more about creators' management of their digital records than we knew about how they managed their analogue materials; we expect to work toward making it possible for creators/donors to self-archive (or archivists to harvest) by establishing trustworthy online transfer, thereby more accurately capturing creators' recordkeeping practices.

Will we continue to use DSpace repository software? This is an open question. We have been using it now for nine years and have worked with it through sequential, evolutionary changes. The latest version includes the ability to delegate control over deposit and descriptive work at different hierarchical levels, and to embargo collections or individual files for specific periods to account for copyright or donor-requested restrictions, changes we welcome to make the work easier. Yet we would like to see the incorporation of more of the processing work into the DSpace platform and may choose to investigate other tools that may provide our students with additional useful experience. In the GCM partnership we are beginning to explore a Fedora repository instance that they have adopted and are paying close attention to the emerging "DSpace with Fedora inside" project of the DuraSpace Foundation.

How can we measure the success of the course thus far? I attempted to do this in 2008.[37] By tracking students who took my digital archiving courses and what they achieved outside and beyond the courses, as well as their post-educational employment, I was able to generalize about outcomes. As to demographics, at that time the courses attracted approximately 50 percent archival studies students, 25 percent preservation administration students, 20 percent

37 The data for the full suite of courses were presented in a workshop, "Education for Digital Stewardship," at the 2008 meeting of the Joint Conference for Digital Libraries.

digital libraries students, and 5 percent information architecture students. Independent student work on digital archives projects (apart from the organized courses) was undertaken by approximately one tenth of all students who have taken courses with digital archiving content. After graduation, of the 115 who could be traced, there were eleven digital archivists (compared to fifty archivists), fourteen digital librarians (compared to seven librarians), three digital records managers (compared to eight records managers), and one post-doctoral digital archives researcher – a total of twenty-nine digital specialists.[38] Anecdotal evidence, including news of students' subsequent employment, indicates that many of the archivists, once on the job, are called on to use the digital skills they acquired; invariably they refer to the hands-on experience they had as giving them the confidence to proceed through reflective practice, and they sometimes get in touch to thank current students for recent project reports. Student evaluations of the digital archiving course have been consistently above four on a five-point scale, and since 2008 the digital archiving course has had to be capped at twenty-five as more students seek it out. Beginning in 2006 I have been fortunate enough to acquire seven doctoral students, all of whom have taken the digital archiving course. Five are specializing in some aspect of digital recordkeeping or archives; all are expected to teach, so it is likely that at least some aspects of the course will be reproduced in the future. This is not enough evidence to assert with confidence that the digital archiving course is meeting all its aspirations precisely, but it does suggest that a follow-up study would be worth carrying out.

Central to the digital archiving problem is incessant change, and one thing worth mentioning is the change that I have seen in students. Their skills are changing, but lest we think that "digital natives" will soon know everything they need to know to do digital archiving, I have disappointing news. Although today's graduate students have been using computers since they were in kindergarten, it is no guarantee that they will know what goes on behind the increasingly seamless interface being provided for consumer electronics. The good news is that today's archival studies students, at least, enjoy the challenge and exotic attraction of legacy computing; we can hope that that will always be the case when today's norm becomes tomorrow's legacy.

Conclusion

Digital materials are mediated by a complex socio-technical infrastructure, one that is constantly evolving and the elements of which are changing at different rates. It will never be possible to know that infrastructure completely, even as it

38 Note that this categorization of positions was based on the job titles of the employing institutions.

applies to only one person through a lifetime of use or even for only one machine used by that person through its much shorter lifetime. The continued expansion of digital inscription as the dominant inscription method in our culture will demand that archivists master its evolving technologies in order to preserve the products of their use for as long as those products are needed, and to support research on those technologies. Aspiring digital archivists must be prepared to know as much about the history of digital technologies as manuscript archivists do about the histories of writing and copying.

Our experience has shown that digital archiving can benefit from a pedagogy that offers students a supportive laboratory experience, mentored by a partnership of experienced faculty, engaged IT staff, working archivists, and local technology experts. This teaching should encourage students to learn to make effective use of reflective practice to query what they know and what the current literature offers, as they engage with real problems and build a stock of documented cases to support the advancement of digital archiving work. We have demonstrated that creating such an environment need not be cost prohibitive. It enables students to move beyond an artificial exercise by engaging with digital materials that are worthy of archival preservation. There is a final advantage to such an approach: students who learn to work in such an environment can potentially replicate it in their eventual places of employment, where it can support the lifelong learning that the archival profession demands.

As Gandalf observed in *Lord of the Rings*, "Perilous to us all are the devices of an art deeper than we possess ourselves"[39] — as archivists of digital inscription, it is up to us to possess and preserve these arts and devices, which have their own cultural value.

39 J.R.R. Tolkien, *Lord of the Rings*, vol. 2: "The Two Towers" (Ballantine Books, New York, 1965), pp. 258–9. This epigraph is especially apposite for archival work in that Gandalf made this observation with reference to the Palantíri, the globes made by ancient art to aid in far-seeing, in time and space.