IFLA 2011
PUERTO RICO
13-18 August, San Juan

World Library and Information Congress: 77th IFLA General Conference and Assembly

Out of the Classroom and into the Laboratory: Teaching Digital Curation Virtually and Experientially

Ross Harvey

and

Jeannette A. Bastian
Graduate School of Library and Information Science, Simmons College, Boston, USA

Meeting: **217 —** *Education for digital curation* **— Education and Training Section with Preservation and Conservation, Information Technology; co-sponsored by ICA Section for Archival Education and Training**

Abstract:

As graduate LIS/IS education seeks to respond to intensifying virtual information and preservation environments, it becomes increasingly clear that innovative teaching tools and methods are required. These teaching tools must complement and enhance state-of-the-art curriculum offerings in subjects such as digital curation. The digital curation courses offered in the LIS School at Simmons College, Boston illustrate an innovative virtual and experiential approach. At the heart of these courses is the Digital Curriculum Laboratory, a virtual archives and preservation laboratory. This paper discusses and demonstrates the relationship between a digital curriculum laboratory, the successful delivery of a digital curation curriculum and its wider international implications.

Introduction

This paper discusses and demonstrates the relationship between a digital curriculum laboratory and the successful delivery of a digital curation curriculum. The first section focuses broadly on pedagogical challenges in a virtual environment, particularly from a cultural heritage informatics perspective; the second describes the Digital Curriculum Laboratory (DCL) currently being developed at the Graduate School of Library and Information Science at Simmons College, Boston; and the third notes both the online and face-to-face digital curation courses in the Simmons program and offers examples of how the DCL is integrated into the digital curation courses through problem-solving scenarios. The final section of the paper deals with the national and international implications of sharing the DCL with the global information community.

Pedagogical Challenges in a Virtual Environment

Educating students to work in digital curation requires students to acquire a broad vision of LIS/IS institutions that looks beyond the silos of traditional information practice towards the convergence of a wide variety of data in both virtual and physical forms. Cultural heritage informatics, a relatively new discipline arising from this convergence, which emphasizes collecting, managing, supporting, reconciling and merging digital data across a broad spectrum of libraries, archives and museums, offers an overarching context supported by digital curation.

Cultural heritage informatics is the general context that the Simmons College digital curation curriculum is focused within and this discussion is offered within that context. Teaching and learning in cultural heritage informatics presents some pedagogic opportunities that challenge the traditional LIS/IS curriculum, which has tended to treat libraries, archives and museums and their distinguishing theories as discrete and separate information and cultural entities. While the attributes of flexibility, forward thinking and critical evaluation are as crucial to understanding rapidly evolving information environments as they ever were, important additions to a set of learning skills are required. Among these are the ability to function comfortably in both digital and physical mediums, to move seamlessly and efficiently between both mediums, to recognize and respect the core differences between information disciplines as well as between the information content itself, and to negotiate the ways in which digital environments can overcome information silos to create a universe of access across institutionalized boundaries.

Cultural heritage informatics workers (including digital curators) must learn to appreciate and balance users' needs for immediate, seamless access to information with an understanding of the context in which information resides. Balancing these user needs with respect for the core theories that ground the various aspects of the heritage the materials come from requires a deep understanding of different disciplines as well as of the digital options for convergence and display of the materials of that heritage. The fundamental principles of cultural heritage convergence should relate to maintaining the balance, so that the very different, but equally relevant, missions of libraries, archives and museums are not lost or subsumed in the desire to bring cultural materials and other information together. Access should not trump core values, but equally, core values should not obfuscate access.

From a practical standpoint, students of cultural heritage informatics (who include digital curators) must learn to respect both the physical and the digital, to manage, value and preserve a wide variety of formats, to identify connections, to evaluate and select systems that suit the needs of their institution best, to appreciate and create relationships among materials, and to imagine and implement the merging of contexts and the provision of access. As the real-world environment for cultural heritage institutions requires ever-expanding competencies, preparing students to enter this environment means helping them to acquire an increasingly complex set of tools.

For educators, the widening gap between physical and digital practice often creates dissonance in the classroom where students are learning virtual skills in physical environments. The fundamental assumption, therefore, of a digital curriculum laboratory is that students need to learn about and experiment with digital tools and materials in digital environments.

A more subtle theoretical assumption is that the added value offered by the digital environment and the convergence of information disciplines not only applies to the ability to access materials but also affects other parameters significantly. Context, presentation and relationships are only some of the factors that potentially expand the way we think about, connect and see materials. While, on the one hand, technology offers a set of tools and competencies that facilitate the institutions' abilities to display materials and the users' options to access them, it also brings new dimensions to the interpretation of the materials themselves. For example, the choice of a content management system becomes a crucial decision in determining presentation and access because different systems display and interpret the same information in different ways; although the image data and the metadata may be the same, the user will see different versions of images, depending on which system an institution has chosen. (This is the basis of the Trust Scenario on the Simmons Digital Curriculum Laboratory (http://calliope.simmons.edu/dcl/lab/scenarios/trust), described later in the paper.) The digital environment, by its very nature, contains interpretive elements that need to be considered as part of the convergence process, which is represented in the following image.

Converging practices in libraries, museums and archives (Source: Curry, 2010)

Need for Innovative Teaching Tools and Methods

As graduate LIS/IS education seeks to respond to rapidly intensifying virtual information and preservation environments, it becomes increasingly clear that innovative teaching tools and methods are required. These new teaching tools must complement and enhance state-of-the-art curriculum offerings in cultural heritage informatics. As well as offering virtual environments for students to experiment in, these tools must be flexible enough to respond to rapid developments in practice and standards.

The Archives and Preservation concentration at Simmons College, Boston is changing its curriculum in response to the new information environments and their requirements, especially the need to offer opportunities for students to experiment. To support this new

curriculum a virtual digital curriculum laboratory was envisioned. Its rationale is to offer experiential learning based on a problem-solving approach. Such laboratories already exist in archives and preservation programs in the U.S. and elsewhere, including at the University of Michigan's School of Information, the University of Arizona's School of Information Resources and Library Science, and the Department of Information Technology and Media at Mid Sweden University.

The digital curation courses offered in the Graduate School of Library and Information Science (GSLIS) at Simmons College illustrate an innovative virtual and experiential approach. At the heart of these courses is the Digital Curriculum Laboratory (DCL) (http://gslis.simmons.edu/dcl/lab), a virtual archives and preservation laboratory developed with grant funding. This organized, open, non-proprietary digital space provides integrated access to digital content, content management tools, standards, curriculum-based scenarios, and a workspace for learning modules tied to class outcomes. The DCL offers its users hands-on experience with digital curation processes and procedures in virtual environments.

The Simmons Digital Curation Curriculum and the DCL

The website of the Graduate School of Library and Information Sciences describes a Preservation Management track (http://www.simmons.edu/gslis/academics/programs/preservation.php), noting that it 'combines hands-on practical experiences with the theoretical underpinnings of preservation and conservation'. Recently this track has been expanded to become a Preservation and Digital Stewardship track, which describes a clearly articulated sequence of courses that allows students to specialize in digital curation. This track has been developed because there is an increasing need for professionals who have the ability to plan, manage and implement practices that ensure the long-term integrity and use of digital resources. This includes information professionals who are charged with digital collections, as well as those working within data-intensive jobs that have curation requirements (for example, scientists, analysts, public officials).

Students in the Preservation and Digital Stewardship track study three courses that provide an overview of the field, both digital and analog, then concentrate on either digital stewardship courses or analog preservation courses. Of course, mixing and matching courses is fully acceptable. To complete the Preservation and Digital Preservation track students study, in addition to the course required of all students in the Simmons LIS masters, three courses strongly recommended for the track (*Introduction to Archival Methods and Services, Preservation Management, Digital Stewardship*) and at least three courses selected from the list specified for each specialization. The list of courses students can study in the Digital Stewardship specialization includes courses of more general applicability (*Database Management, Web Development and Information Architecture, Digital Libraries, XML (eXtensible Markup Language), Metadata*), as well as several specifically about digital curation (*Archiving & Preserving Digital Materials,*

Digital Project Management, an independent study on a digital stewardship project, and an internship in digital stewardship). Several of the digital stewardship and digital curation courses use the Digital Curriculum Laboratory.

The formal description of Simmons College's Digital Curriculum Laboratory (DCL) is:

A web-based space providing integrated access to content, tools, curriculum-based scenarios and workspaces. It can be used to experiment with a range of archival and preservation procedures for digital records, following a continuum from record creation through arrangement, description and delivery (http://calliope.simmons.edu/dcl/lab).

The features to emphasize are that it is **web-based**, that it provides **integrated access**, and is available for **experiment** in the context of archives and preservation. The larger space in which the DCL is located is a Cultural Heritage Informatics (CHI) initiative, which is developing new curriculum based on the concept of convergence of practice in a digital world. The CHI web site explains that

Cultural heritage informatics is the study of and the creation of added cultural value by the linking of disparate digital data sets, stored either locally or remotely according to accepted standards of description, arrangement, and metadata for archives, records management, museums or cultural materials. Cultural heritage informatics appraises data and data sets for enduring value in the context of archives or cultural heritage. Cultural heritage informatics explores the creation of new relationships and new knowledge by bringing digital data sets representing social and cultural activity together in novel ways (http://calliope.simmons.edu/dcl/culturalheritage).

The DCL has been developing since 2008, when several Simmons faculty recognized the need to provide hands-on experience with digital records and archiving. They received a small Simmons internal grant (Curriculum Technology Support Grant awarded by the Pottruck Technology Resource Center of Simmons College: $2,500) to carry out initial scoping of the laboratory. This led to successful grants from the IMLS (Grant Number 113 2435 20 400129 ($455,000) - includes funding for the DCL) and NHPRC (Project Name: Archives and Preservation Digital Curriculum Lab ($138,000) - specifically to build the DCL).

The DCL can be envisaged as two complementary parts. Sitting in the DCL virtual space are:
- Scenarios: the problem statement that drives the choice of the other components of the DCL
- Applications: called on as required to address the problem
- Exercises: intended to equip DCL users with appropriate skills
- Content: used as required by the scenario.

Sitting outside the DCL, but required for its use, are:
- Applications
- Standards
- Learning Management System(s)
- Groupware.

The DCL does not provide everything that is needed to address each scenario. The computers used to access the DCL through the web need to have some applications loaded and also need to provide access to web-based standards, an LMS (Learning Management System, such as Blackboard or Moodle) and groupware (for example, Google Docs) as required by the scenario. Each scenario or exercise in the DCL specifies what is required. For example, one preservation scenario requires Microsoft Word and Open Office to be available on the client PC, and, when used in the Simmons context, also requires access to the LMS (WebCT/Vista); the scenario itself and the content needed to test the scenario are available from the DCL.
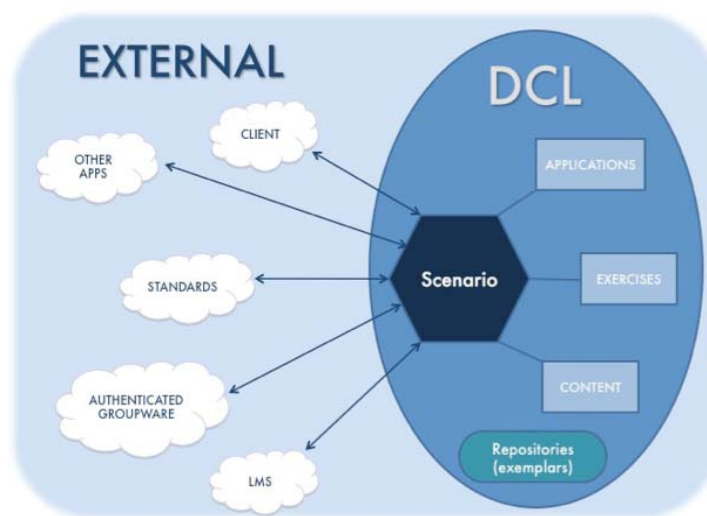
Cloud model of DCL (Source: http://calliope.simmons.edu/dcl/lab)

The 'cloud' model of the DCL illustrates these two parts and indicates the current thinking of the team developing the DCL. This model will continue to evolve as the DCL expands.

Digital Curation Courses at Simmons

Two of the courses offered at Simmons are next noted in order to illustrate the use of the DCL in teaching digital curation.

LIS 444 *Archiving and Preserving Digital Media* has been available face-to-face for some years and from 2011 is also offered as a fully online course. It examines the challenges and issues of archiving and preserving digital media and notes the practices that have been developed (and are still evolving) to archive and preserve digital materials. Students become familiar with relevant skills by working on a range of practical exercises and on developing a digital archive.

LIS 531W *Digital Stewardship* was offered face-to-face for the first time in 2011 as part of a new curriculum in Cultural Heritage Informatics. Its focus is more general than the focus of LIS 444, concentrating on broader policy issues and the implications for practice in cultural heritage organizations. It is particularly concerned with the long-term sustainability of digital repositories, libraries and archives. This course is not currently offered online, but will be redeveloped to be available in this mode in the near future. Inevitably there is some overlap between these two courses. The intention is that LIS 531W *Digital Stewardship* becomes a more general overview course and LIS 444 *Archiving and Preserving Digital Media* is redeveloped as a course that focuses on the IT aspects of digital curation. Both courses will provide students with practical skills, although these will be considerably more advanced in LIS 444.

Using the DCL

The DCL is currently used in both LIS 444 and LIS 531W to provide access to scenarios, exercises and standards. The example of the online version of LIS 444 best illustrates how the DCL is used in a virtual learning environment. The DCL provides links to exercises, scenarios, applications and content used in the LIS444 classes. When students log in to the DCL using login details specific to this course, they see this view of the DCL:

(Source: http://calliope.simmons.edu/dcl/lab/portal/access)

This page provides access to the scenarios and exercises students use in the LIS 444 course. The exercises and scenarios are intrinsic to the course: as indicated earlier, the scenarios provide the problem statement that drives the choice of other contents of the DCL; applications are called on as required to address the problem, as is relevant content; and exercises equip DCL users with appropriate skills that are needed to address the problems posed by the scenarios.

Nine preservation-related exercises are currently available on the DCL to students in LIS 444. They aim to provide some understanding of common digital preservation techniques and tools.

The exercises are:

1. Migrating and verifying files
2. Encapsulation
3. Assessing Robustness of File Formats
4. Emulators
5. Building a Web Archive
6. Digital Preservation Tools and Services
7. Preserving Digital Personal Files
8. Making a Video about Digital Preservation
9. Policies for Digital Preservation.

Each exercise is laid out in a standard format: a brief description of the aim and requirements; a list of computing and content requirements for the exercise; a scenario if relevant to the exercise; and the tasks required of the student. There may be links to external websites, for example, where software can be downloaded or where further information is available. The intention is that students can use the lab with a minimum of instruction (preferably none) beyond that provided in the exercise. Exercise 2 illustrates this standard layout:

Exercise 2: Encapsulation

This exercise requires you to install the Xena application and use it to create Xena information packages.

Computing Requirements

To complete this exercise you need:

- A computer connected to the Internet (either Mac or Windows OS)
- Open Office software installed
- Administrator rights for the computer so you can install Xena software and other software that Xena requires

Your Tasks

1. Read carefully the documentation about Xena so that you understand what it aims to do. Start with the FAQs (http://xena.sourceforge.net/faq.php) and then check out the Xena wiki (http://sourceforge.net/apps/mediawiki/xena/index.php?title=Main_Page). Be sure to look carefully at the sections "Installing Xena and "Configuring Xena"
2. Make sure you know what outputs you expect Xena to produce
3. Download and install the Xena digital preservation software and relevant plug-ins (http://xena.sourceforge.net/). (A reminder: you may also need to install Open Office (http://www.openoffice.org/) which is required for Xena to work)
4. Select five files that Xena can handle (the file types that Xena accepts are identified in the documentation). The files should be of different file types. Process these files using Xena
5. Submit the Xena files you have created (or provide a link to them where they can be accessed) and a brief report on how Xena is useful for digital preservation purposes.

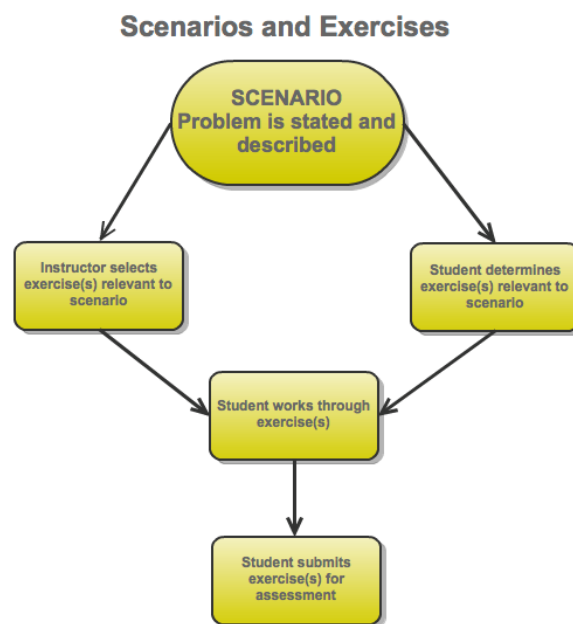The exercises can be used as stand-alone exercises or can be linked with a scenario. One of the scenarios used is as follows:

You are in charge of the preservation program of a university library. A former faculty member of the university has offered his papers to the university library, and this offer has been accepted. The donation includes a number of three-and-a quarter inch diskettes and Zip

disks containing digital files that need to be preserved so that they are usable in the future. The files are all word-processed files, files created using simple database software, and files created by Microsoft Word. They date from the mid 1980s to about 2005.

Several exercises are relevant to this scenario. Exercise 7 provides skills in identifying old file formats, exercise 3 in determining which file formats are preferred for long-term preservation, exercise 1 in running checksum software and observing the consequences of migrating files, and exercise 2 in using Xena software to develop information packages.

The way in which these preservation exercises are used is illustrated in the following diagram. Instructors can advise students which exercises are relevant to a scenario, or alternatively students can experiment:

Scenarios and Exercises

SCENARIO
Problem is stated and described

Instructor selects exercise(s) relevant to scenario

Student determines exercise(s) relevant to scenario

Student works through exercise(s)

Student submits exercise(s) for assessment

Each exercise calls on other materials that may be located in the DCL or elsewhere, as indicated in the following diagram. These materials can be content (files that the student applies the exercise to), standards relevant to the exercise or specified content, and software applications needed for the exercise. The outcome of the exercise can be in the form of a written report and/or files, or another product as specified by the exercise.

Preservation Exercises
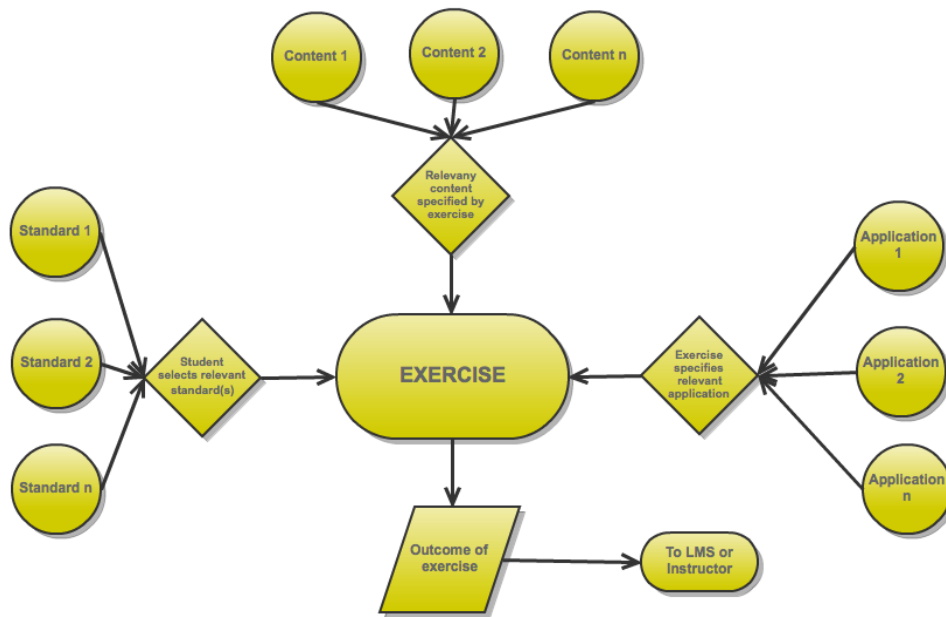
LIS 444 students are also required to participate, for assessment, in the development and maintenance of a class digital archive, which uses the DCL, based on the theme 'your time at Simmons College'. This was started in 2010 and has resulted in an archive populated with some content. Students are asked to consider the characteristics of a digital archive in general 'big picture' terms, to select files to contribute to the archive (four or five on the theme 'your time at Simmons College') and to identify the actions required before the files can be added to the archive (such as adding metadata, checksums, virus checking). The DCL is used for the next actions. Students identify a suitable software application for long-term digital preservation in which to store the files selected and the metadata they have added, selecting the application from the open source applications currently installed in the DCL (Alfresco, Collective Access, DSpace, ePrints, Fedora, Greenstone 3, Omeka). The DCL's *Preservation Scenario 2* (http://calliope.simmons.edu/dcl/lab/scenarios/preservation2) provides a focus for this. The students are then advised that they will be storing their files in an ePrints repository, of which an instance is mounted on the DCL, and they then upload their files and metadata to ePrints.

LIS444 Class Archive

Items where Year is 2010

Up a level

Export as [ASCII Citation ▼] [Export]   RSS 2.0 RSS 1.0 Atom

Group by: **Creators** | Item Type | No Grouping

Jump to: D | G | I | J | K | L | M | S | W

Number of items: **35**.

D

DeClue, J (2010) *Bibliography.* Other. DeClue Juliette. (Submitted)

G

Griffin, Erin (2010) *Deed of Gift.* [Teaching Resource] (Unpublished)

I

Ingrao , Katherine (2010) *Codzilla Ride.* [Image] (Unpublished)

Ingrao , Katherine (2010) *Reference for Seniors Presentation.* Microsoft Office Suite. (Unpublished)

Ingrao , Katherine (2010) *Touring D.C. During ALA 2010.* [Image] (Unpublished)

J

Jorgensen, Anne (2010) *Boston Public Garden.* [Image] (Unpublished)

(Source: http://liseprints.gslis.simmons.edu/)

As the final part of this assignment students are asked to determine what else needs to be in place to ensure that this ePrints archive is sustainable over the next twenty years. Future classes will be involved in additional activities, such as adding files and metadata created by students in earlier classes but not yet uploaded to the archive, developing policies for this archive, editing the existing metadata to confirm to the policies, improving the existing metadata, and determining how to make the ePrints archive function as a long-term preservation tool.

The DCL is still being developed for digital curation teaching. Additional practical exercises have been developed by a doctoral student, Patricia Condon, and will be added to the DCL in the near future. They include:

1.   Automating part of a workflow - this uses readily available automation tools such as Automator for Mac OSX.
2.   Working with Submission Information Packages (SIP) - using tarballs
3.   Developing a template/workflow for archiving and curating personal digital files using commonly available free software

4.   Creating manifest files for digital transfers

5.   Automatically generating metadata using the NZ Metadata Extractor.

Conclusion: International Implications

As a pilot project in international cooperation, the DCL is being developed in conjunction with the Department of Information Technology and Media at Mid Sweden

12

University in Härnösand, Sweden and University College, London. The partnership is still in its early stages but already offers potentially useful and exciting outcomes. The general and universal applicability of digital curation education suggests the easy and fruitful interchange of courses. Group projects with students from all three countries working together in the DCL environment will not only give students exposure to other ways of thinking and working but will also prepare students to work in the international arena that is increasingly central to digital curation practice. Our reason for using only open source software is to be able to offer the DCL for use in LIS/IS education in its widest possible context, since we feel that the tools it contains are generally applicable in any digital environment.

From the cultural heritage informatics standpoint, our partnership offers creative opportunities for teachers and for students, not only in teaching and learning about different culture heritages, but also in direct interaction with differences and similarities across institutional and global environments. In the convergence across national borders, we see digital curation as offering trans-national opportunities for the sort of cultural understanding and recognition that educators are only beginning to recognize and to teach.

References

Anderson, K., Bastian, J., Harvey, R., Samuelson, G., Plum, T. (forthcoming) 'Teaching to Trust: How a Virtual Archives Curriculum Laboratory creates a Global Education Community', *Archival Science*.

Bastian, J., Cloonan, M., Harvey, R. (2011) 'From Teacher to Learner to User: Developing a Digital Stewardship Pedagogy', *Library Trends*, v59 no4: 607-622.

Bastian, J., Harvey, R., Mahard, M., Plum, T. (2010). 'Building a Virtual Archives and Preservation Curriculum Laboratory at Simmons College: A Case Study in Collaborative Construction', *JELIS: Journal of Education for Library & Information Science*, v51 no4: 241-250.

Curry, D. (2010). 'Thinking about Convergence. Part 1. Drivers', *Future of Museums* blog, March (futureofmuseums.blogspot.com/2010/03/one-potential-future-for-museums.html).

Contact information:

Dr Ross Harvey, GSLIS, Simmons College, 300 The Fenway, Boston MA 02115, USA. email: ross.harvey@simmons.edu)

Dr Jeannette Bastian, Professor, GSLIS, Simmons College, 300 The Fenway, Boston MA 02115, USA. email: Jeannette.bastian@simmons.edu)

Biographical information:

Ross Harvey is Visiting Professor in the Graduate School of Library and Information Science, Simmons College, Boston, a position he has held since 2008. Before joining Simmons he was the inaugural Professor of Library and Information Studies at Charles Sturt University, Australia, from 1999 to 2008, and he has held positions at other universities in Australia, Singapore, and New Zealand. Visiting Professorships at the University of British Columbia (2008) and the University of Glasgow (2007-2008) allowed him to observe first hand current digital preservation practice and participate in digital curation research. Harvey's research and teaching interests focus on the stewardship of digital materials in libraries and archives, particularly on its preservation, and on the history of the book. He has extensive experience in research projects in Australia and the U.K., most recently with the Humanities Advanced Technology and Information Institute at the University of Glasgow and the Digital Curation Centre. He has published widely in the fields of bibliographic organization, library education, the preservation of library and archival material, and newspaper history. His latest book is *Digital Curation* (Neal-Schuman, 2010).

Jeannette A. Bastian is a Professor at the Graduate School of Library and Information Science, Simmons College, Boston, Massachusetts where she directs their archives education program. During Spring 2011 she is Acting Dean of GSLIS. She was the Territorial Librarian and Archivist of the United States Virgin Islands from 1987 to 1998 and received her Ph.D. from the University of Pittsburgh in 1999. Her research interests and writings are in the areas of post-colonialism, collective memory and archival education. Her publications include *West Indian Literature: An Index to Criticism, 1930-1975* (writing as J. Allis, 1981), *Owning Memory: How a Caribbean Community Lost Its Archives and Found Its History* (2003), *Archival Internships: A Guide for Faculty, Supervisors, and Students* (2008 with Donna Webber) and an edited volume, with Ben Alexander, *Community Archives: The Shaping of Memory* (2009).