

Digitization in the Real World

**Lessons Learned from
Small and Medium-Sized
Digitization Projects**

Edited by

Kwong Bor Ng & Jason Kucsma



Metropolitan New York Library Council

Published in the United States of America by
Metropolitan New York Library Council
57 East 11th Street, 4th floor
New York, NY 10003-4605
p: (212) 228-2320 f: (212) 228-2598
Web site: <http://www.metro.org>

ISBN: 978-0-615-379998-2

Cover Design: Jason Kucsma (*illustration by Smartone Design,
licensed via iStockphoto.com*)

Reviewers Committee: Mark F. Anderson, Jill Annitto, Anna Craft, Jody DeRidder, Renate Evers, Wei Fang, Maureen M. Knapp, Sue Kunda, Mandy Mastrovita, Ken Middleton, Emily Pfotenhauer, Mark Phillipson, Alice Platt, Mary Z. Rose, Stacy Schiff, Jennifer Weintraub, Andrew Weiss.

Copyright © 2010 by Metropolitan New York Library Council. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

The views expressed in this book are those of the authors, but not necessarily those of the publisher.

METRO Grant Success Story: Waterways of New York Project

Claudia A. Perry and Thomas T. Surprenant
(Queens College, CUNY.)

Abstract

The concept of experiential learning is particularly useful when students are required to create database entries as part of an ongoing, real-life, online experience. A METRO grant resulted in an opportunity to use students to create a CONTENTdm database which, with the continued software support from METRO, has continued and evolved until the present. This chapter describes the experience of both faculty and students. Sections include the background, technical issues and implications for teaching, project procedures and workflow, successes and lessons learned, challenges and next steps. Of particular interest is the use of out of copyright postcards and the metadata that has resulted from intensive student study and evaluation of the data contained on these cards. Those contemplating a digitization project of their own will be able to learn much about best practices, project planning, management and the advantages/disadvantages of the CONTENTdm software.

Keyword: Best Practices, Canals, Case Studies, Cooperative Learning, Digitization, Digital Collection Management Software, Digital Collections, Digital Imaging, Experiential Learning, Library Education, Metadata, Postcards, Project Based Learning, Project Management, Project Planning, Quality Control, Standards, Student Developed Materials, Student Participation, Student Projects, Waterways.

Introduction

For many of us, hands-on learning is the best way to integrate an understanding of principles and best practices with a practical grasp of the actual challenges and learning opportunities of a project. This is particularly true for library school graduate students seeking to expand their theoretical, technical and management skills. As digitization is increasingly seen as a worthy endeavor for even the smallest institutions, it is worth considering the range of approaches available for gaining needed expertise, especially at the novice level. Examining the long-term development of an integrated, semester-long, course-based approach to digitization may be of value for those seeking an inexpensive approach for the creation of small to medium-sized digital collections.

A course entitled “Introduction to Digital Imaging” was first taught at the Queens College Graduate School of Library & Information Studies (GSLIS), City University of New York (CUNY), in the Fall of 2003. In the Spring of 2005, a year-long METRO-funded grant facilitated a co-operative project between the Rosenthal Library and the GSLIS to support student digitization of a portion of the Queens College Rosenthal Library Archives (e.g. see GSLIS, 2005-2009, Digitization projects). The project included a variety of forms and formats. The evaluation of this valuable learning experience identified a strong need to find a single standard format that was information rich and moderate in scope, but which lent itself to more uniform metadata standards and digital specifications. The evolving project, “Waterways of New York”, an online digital collection of historical postcards, was created in 2006, and partially supported by METRO through continued access to CONTENTdm. It continues to be extended by GSLIS students each semester the course is taught.

Scope and Format

The most important feedback provided to our team by METRO digitization experts regarding our “Rosenthal Library Archives” initiative was the value of working with a limited number of manageable formats and a relatively focused subject area and time

frame. During the implementation of the grant a serious problem was the complexity resulting from too many different types of media, the overly wide range of subject matter, and the challenges these characteristics presented to the creation of consistent metadata.

One of the GSLIS professors, Thomas Surprenant, has an ever-expanding collection of Erie Canal and related New York State waterways antique postcards, which addressed many of the problems noted in the METRO feedback. In particular, by selecting a single, simple, information rich format—postcards published before 1923—copyright concerns were eliminated and only a single set of digitization specifications needed to be developed. METRO’s willingness to host the collection on their CONTENTdm server simplified selection of Dublin Core as the metadata standard, and use of a subset of the Library of Congress Thesaurus of Graphic Materials (TGM) for standardized metadata terminology (Library of Congress, 2007). This greatly aided our ability to develop a manageable set of project-specific guidelines that could be adequately addressed by the evolving documentation.

The choice of postcards as the source medium turned out to be far more interesting to the students than was expected. An initial option to describe the backs—as well as the front images of cards—was enthusiastically embraced by virtually all of the students and became the norm for subsequent classes. Hand-written messages, address conventions, postmarks, trademarks and other attributes of the cards were at times as much or even more rewarding to analyze than the front images themselves. Further, student interest in the varied aspects of architecture and activities of daily living portrayed in the postcards led to an expansion of emphasis far beyond the initial focus of the project on locks, canal boats, shipping, waterways and transportation.

Background, Technical issues and Implications for Teaching

Any planning for digitization requires a detailed analysis of one’s institution, and an assessment of where the proposed project fits into

its mission and priorities. Further, consideration of the potential audience(s), project goals and objectives, resources and limitations, oversight and long-term maintenance are among the many issues to be addressed (e.g. see JISC Digital Media, 2008: Project management; North Carolina Echo Project, 2007). These considerations inevitably will shape the nature of the evolving project. It is important that an honest appraisal be conducted, committed to writing, and approved by the appropriate governing bodies. However, the nature of digital projects ensures that adjustments inevitably will be required over time. Changing standards, software and hardware upgrades, technical glitches, and shifts in the growth of a project are just a few of the issues which must be dealt with, often on very short notice. Planning and documentation therefore should be viewed as an iterative process, where ongoing evaluation is used to address and correct for changing circumstances.

Creating a list of stakeholders and intimately involving them in this planning process is critical to success. In our own case, student feedback on procedures and emphasis has been an invaluable aspect of the evolving project. Each incoming class section serves as a de facto Advisory/ Editorial Board that contributes to the decision-making process. These contributions include identification of additional TGM terms for our thesaurus, the development of standardized Trademark descriptions, fine-tuning of documentation and lab handouts, and increasingly higher expectations for the quality of the metadata. Within a more traditional library environment, all members of the digitization team, as well as users and other staff members, undoubtedly will have many valuable insights to contribute.

Among the key elements shaping the evolution of a project-based digitization course at Queens College were the following:

Institutional characteristics

- When the initial course was developed it was necessary to have the course proposal cleared with the departmental Curriculum Committee after consultation with the Chair. This required the development of course goals and objectives, specific readings, and course assignments and activities.

- After three semesters teaching the course it was submitted to the GSLIS and College Graduate Curriculum Committees, Faculty Senate and, ultimately, the CUNY Board of Trustees for approval as a permanent course.
- An understanding of the possible pitfalls of the process at every step was important to ensure that all potential hurdles were considered and cleared.
- Even outside of explicitly academic environments, proper attention to obtaining documented approvals and support from key stakeholders--at all levels up to the governing board--will prove invaluable in avoiding challenges and ensuring continued buy-in by the institution and other funding agencies.

Lab facilities (capabilities and challenges)

- For our project we were able to use a 16 workstation Mac lab with direct connections to the Internet. The lab had been expressly designed by GSLIS faculty for digitization-related activities and hands-on learning, in close collaboration with the Queens College Office of Converging Technologies (OCT) and college architectural staff, in conjunction with the development of the course proposal. Appropriate institutional commitment to fund, support and regularly upgrade such a lab was, and is, essential to the continuing success of the project.
- Specifications included an instructor workstation (in addition to student Macs), ceiling mounted projector and wall screen for demonstrations, two (eventually three) flatbed scanners, SilverFast AI scanning software, Photoshop, and the Microsoft Office Suite, particularly Excel.
- A major continuing challenge concerns computer and software upgrades. The OCT staff do not always consider the rhythms of the academic year in making changes to the lab, which regularly causes problems, even after many years of teaching the course. For example, in the Fall 2009 semester alone, new computers were installed during the first week of the semester. This resulted in equipment and software glitches, and a delay in the availability

of the lab, as well as the need to test software functionality and then revise/upgrade lab handouts with minimal advance notice.

- While we were grateful for the new equipment (a regular replacement cycle is essential for ongoing functionality), timing issues resulted in a rough start to the semester.
- A major equipment problem for us was solved when Apple changed to an Intel CPU. The new Mac computers are now dual boot (Apple and PC Operating Systems), meaning that they can now run CONTENTdm (CDM) using the Project Client software interface. Previously, lab sessions had to be specially scheduled in a nearby PC lab (CDM Project Client software is not available for the Mac OS). However, dual boot capabilities have created additional problems of compatibility, accessibility and ongoing troubleshooting.

Software

- As noted above, new equipment means software installation and the attendant complications. The specialized nature of our lab, and lack of teaching assistants, necessitates that course faculty test all functionalities and work with OCT staff to address problems. Oftentimes this has meant repeated testing and troubleshooting, frequently a day or two prior to a scheduled class. Such technical malfunctions can wreak havoc on the best-planned teaching schedule.
- While CDM has been sufficient for our needs, and we are extremely grateful to METRO for their continuing support, there are still some issues that cause concern. The biggest is that students cannot directly upload their input into the database due to the administrative rights structure. This situation requires another level of review by the course instructors serving as database administrators/quality control experts, adding substantially to time demands near the end of the semester. In addition, after students submit their data entries for approval, editing ability on their side is extremely constrained, by both time and software limitations.

- More recent upgrades appear to have adjusted this limitation, permitting downloading of materials from the live database for additional editing if errors are detected. However, this creates additional levels of oversight and complexity, and assumes that the instructors will be able to approve the uploads in time for the students to review and make changes. This is simply not readily accommodated within a 15 week course schedule.
- Further, although recent versions have been more stable, in the past CONTENTdm has crashed frequently, causing much frustration on the part of both students and faculty.
- These points emphasize the steep, and ongoing, learning curve of digital project-based courses for faculty, support staff and students.

Support

- Adequate and timely support for equipment and software is essential to any technology-based project. The GSLIS has a number of student computer assistants and a campus-wide Help Desk, but as noted, the specialized nature of our lab sometimes puts it outside of the realm of their expertise.
- It is good practice to fully document and save ALL help desk requests and related support communication. These include emails, screen shots and help desk tickets. These records of ongoing and recurring problems have proved to be invaluable in our efforts to ensure follow-through, and to support our case when requests have not been fully resolved to our satisfaction or when problems repeat themselves.
- CONTENTdm Help seems to work best when we go through METRO. That means that an additional layer of contact needs to be activated anytime is a problem. That said, all relevant staff at METRO, over the years of this project, have been incredibly knowledgeable, supportive and responsive to our needs.

Staffing/oversight

- Experience suggests that having a subject expert for image content is a critical factor. The faculty have, or have developed over time, sufficient expertise to assist students in their metadata and description activities.
- Given the need to protect the postcards and the equipment the lab has to be under supervision whenever anyone is working. This greatly adds to the time burden of both faculty and staff.
- Postcards are stored in archival quality sleeves and students use white gloves when handling the postcards while scanning.

Class size and student characteristics

- The class size is dictated by two elements: the number of work stations and the work volume. Experience suggests that all students need to have access to their own workstations, and two workstations are dedicated to scanning use (a third must be shared between functions). The initial classes scanned, created metadata and submitted for approval six cards (front and back), but many of these initial canal cards were fairly simple rural scenes. Given the amount of detail that has emerged in postcards in later semesters, we have gradually reduced input to three cards (front and back), because the quality and quantity of the metadata has increased substantially. The time spent in quality control by instructors has increased commensurably, despite having students doing quality control on their partners' work.
- Those involved in the digitization process are best served with at least an intermediate level of computer and software expertise. We have constructed our teaching labs in a step-by-step fashion, and utilize in-class time extensively. This allows the faculty to introduce and demonstrate skills and to detail the various steps of the project.
- In our experience, students who are highly competent in computers and/or relevant software or metadata creation have been more than eager to assist their classmates. This leads to a

highly supportive class environment in which all learn from one another, modeling (one hopes) the ideal workplace environment.

- However, a fair number of students have no previous familiarity with the Mac OS, Photoshop, scanning and related software, which complicates the pace at which the class can proceed. The nature of our curriculum and scheduling constraints make it difficult to require pre-requisites beyond the required core courses. Consequently, a teaching assistant to help in quality control, and in the provision of additional technical support in lab sessions, would be extremely desirable for all.

Evolving nature of the target collection

- Initially, the Waterways Post Card Collection consisted principally of cards of the Erie Canal, with collections of additional New York State canals such as the Oswego, Seneca, and Champlain Canals. As demand for the course has remained steady, indeed increased, the diminishing availability of canal-related cards posed a potential problem. (Most cards are currently obtained on eBay, and changing availability in geographic scope is an interesting topic for another paper.) On the other hand, many of those canal-related cards that have become available are increasingly distinctive.
- With the Quadricentennial Celebration of the discovery of the Hudson River approaching in Fall 2009, it appeared to be a logical extension of our collecting scope to extend to another key New York State waterway, the Hudson River. We included in our selection criteria cards depicting New York Harbor and the East River. The first such Hudson River cards were digitized in Fall 2008.
- The expansion in scope of the cards created fascinating but unanticipated challenges. Publishers, trademarks, and the increasing complexity of the images depicted required a substantial expansion of the TGM thesaurus, as well as the development of descriptions of an increasingly diverse set of trademarks, logos, stamp boxes and postmarks.

Consistency and accuracy

- Digitizing any collection over a period of time by a changing group of participants creates somewhat greater consistency and accuracy problems than might apply in a short-term. In spite of the iterative editing of documentation, inconsistencies and errors are regularly emerging in our project. To a large degree, this is due to the pressures of a fast-paced curriculum, a constantly changing panoply of operating systems, software versions, additions to our thesaurus and the ongoing, changing nature of the cards within our purview..
- Our experience has shown us that the students themselves are the best editors in catching errors and inconsistencies. It is obvious that the road to “perfect” metadata, documentation, labs and handouts is continuous, difficult, and perhaps, ultimately elusive. Such is the nature of a work in progress.
- It helps to have students who have a keen eye for detail as well. The project was significantly enhanced when then-student Susan Savage completed an Independent Study project in Spring 2007, that corrected many of our past mistakes, and developed the scaffolding of our current metadata documentation. It is now obvious that outside help in editing is an important part of the process (although not easily achieved).

Key readings and course activities and relevance for the project

- Clearly, carefully selected readings are critical to the success of a digitization project. We provide access to a range of resources in an effort to meet the needs of those at varying levels of familiarity with digitization and related issues. Alumni feedback has suggested the importance to many of providing continued access to the resources once our graduates are working in the field. Once involved in a real world project, many become even more aware of the importance of items that may not have seemed salient at the time of the course.
- Students complete a “Tech Review Exercise” in week seven, to document their understanding of key technical concepts. This

reduces the need to cover many basic concepts during lectures, and to focus on practical development of skills and discussion.

- In addition, students prepare a detailed case study of a well-documented digital collection, and present an overview of their sites to class members. Examining the successes and challenges of a substantive collection reinforces the concepts and principles learned and applied throughout the course. Students comment with great pride on the degree of detail of the metadata they have developed for their cards, as compared to some of the case studies examined, including many well-funded projects. For many respondents to our end-of-semester evaluation surveys, the case study is perceived as a high point of the course.

Project Procedures and Workflow

Assignment of cards to students

- Postcards are carefully selected in order to give each student a maximum of exposure to bibliographic richness, varying formats and levels of difficulty. For instance, at least one card will have writing on it for transcription. Cards also are selected on the basis of color or black & white, as well as postcard era (e.g. divided back; undivided back). In addition, postmarks, publishing information, trademarks and stamps are considered in allocating cards. A lecture on the history of postcards provides background in understanding the evolution of these standards (e.g. see Smithsonian Institution, n.d.).
- Once selected, each card is given an eight alphanumeric accession number that identifies the subcollection, card number (a total of 9999 entries are available) in that series, the type of image (access or archival), and whether or not it is a front or back view. Early on in the project it was realized that having a related but unique identifier for each side was vital to having both sides of a card displayed together. At this point the Waterways series identifiers are: Champlain (c), Erie (e), Delaware and Hudson (d), Hudson River (h), and Seneca (s). Thus, a card in the Hudson River series

available on the Web will have two accession numbers – one for the front and one for the back: e.g. h0062ac1, h0062ac2. Slightly different filing naming schemes are employed for the master archival files (e.g. h0062ar1.tif), so that the nature of the file format is evident even without the file extension.

- We adhere to the ISO 9660 8.3 alphanumeric naming standard to ensure that our files will be compatible across platforms (JISC, 2008, Choosing a file name).
- Stickers with the student names and the accession numbers are printed and attached to the outside of the archival sleeves that contain the assigned postcards. The sleeves are put into a binder so that students can access them, and photocopies of fronts and backs created to ensure a record of assignments.

Scanning and Creating Derivatives

- Once the postcards are distributed to students, they engage in hands-on instructional labs in scanning. Students use SilverFast AI software, in conjunction with the latest version of Photoshop, using the currently available flatbed scanners (these are regularly updated) to create uncompressed Tagged Image File Format (TIFF) scans. The resulting archival files, in the 20Mb range, are stored (and backed up) off line and can be examined by students if extensive detail is needed. These files are the heart of any database because they can be used for non-web purposes
- Using Photoshop, students then create JPEG derivatives at medium compression with a resolution of 150 ppi with 1000 pixels on the long dimension (following the *CDL Guidelines for Digital Images*: California Digital Library, 2009). The “ar” on the archival files are changed to “ac” on the accession numbers to reflect the change in file size and specifications. In this process the archival masters are reduced in size to approximately 200 KB, appropriate for web viewing, and yet still large enough to zoom for some degree of detail.

Assigning Metadata

- The heart of the course is when students assign metadata. A group exercise using different sample postcards for each pair of students, which is then discussed within the class, provides basic experience in assigning subject terms. What is initially thought of as a simple process quickly becomes complex when the students are confronted with the reality of their individual cards.
- Each student is given a thesaurus, based primarily on the Thesaurus of Graphic Materials (TGM), that contains all of the subject terms to be used; an electronic version also can be used to copy/paste the terms. In addition, detailed handouts with trademark, stamp data, and metadata procedural guidelines are made available in print and digital form. Metadata and JPEG images will be inserted into an Excel template, which lists all of the metadata requirements according to CDM fields (tailored to our project), and the corresponding Dublin Core fields.
- As needed, additional terms from the TGM are added to our project thesaurus to reflect characteristics not previously encountered. For example, the move to using cards depicting New York Harbor required the addition of such terms as skyscrapers, aquariums and ferries.
- Each week, short lectures introduce the students to canal history, lore and terminology. Once the basics are covered students are then given time to insert JPEGs of the fronts and backs of their first card (the simplest) into an Excel template. They then proceed to select subject terms, develop descriptions, and identify key Dublin Core fields (e.g. Title, Creator, etc.).
- The faculty provides assistance and guidance during this process, and students are encouraged to work with a partner. Initially there is a great deal of trepidation, but as students gain experience and confidence less faculty attention becomes necessary.

Quality control teams

- In order to provide more experience to the students, and to serve as a double check before the metadata is passed on to the faculty, they are paired up with a quality control partner, and required to exchange their Excel metadata files. Each student then reviews his/her partner's metadata, starting with the initial card, and makes any necessary editorial changes and/or suggestions for improvement using a different colored font. The reviewed metadata set is then forwarded to the faculty for their comments and edits, added in yet a different color font. This provides an iterative record of changes: the input of initial author and control partner, and final corrections by faculty. After receipt of faculty feedback, the teams can then proceed with subsequent cards.
- This staggered approach is recommended because students tend to make the most errors in their first attempts, and to learn from the ongoing feedback. Additional benefits to the project include the identification of new trademarks in the initial set of cards, descriptions of which can then be shared with others on subsequent cards. Lastly, repeated mistakes may reveal unintended errors or inconsistencies in documentation that can then be revised to reduce future errors.
- While this would seem to be a straightforward process it has turned out to be much more challenging than first envisioned. Yes, the students get, essentially, experience with six, not three cards which is good. But a number of other problems have resulted. It is almost always the case that the better students pair up and that they turn in metadata that needs little comment. However, it still takes an average of fifteen minutes per relatively error-free file for faculty to review submissions. Weaker students require much more faculty review time either because the editing process is sloppy or submissions are late. In some cases faculty review and editing can take as much as an hour per Excel file. Given time constraints and course deadlines this can get extremely stressful to all parties.

- That said, it continues to surprise the faculty that: 1) the quality of most metadata submissions is so impressive, and that 2) there remain previously unidentified errors in what seems to be a fairly strict process. In this regard the quality control process is working as envisioned.

Below is a transcribed example of the metadata for “1609 • HUDSON-FULTON CELEBRATION • 1909 [front caption] (1front) [h0189ac1]” after it has been uploaded to CDM.

Title: 1609 HUDSON-FULTON CELEBRATION 1909 [front caption] (1front) [h0189ac1]

Creator: Copyright 1909 J. Koehler, N.Y. [indicated on front only]

Subject—Front: Cliffs, Clouds, Flags, Grasses, Portraits, Rocks, Ropes, Schedules (Time plans), Ship equipment & rigging, Shrubs, Smoke, Smokestacks, Steam engines, Trees, Men, Passengers, People, Color postcards, Sailing ships, Side wheelers, Aerial views , Rivers.

Description—Front: A commemorative postcard celebrating the 300th anniversary of the discovery of the Hudson River, with portraits of Henry Hudson and Robert Fulton superimposed over a daytime aerial view of the Hudson River. Prominently featured are a sailing ship (circa 1609) and steamship (circa 1809) [presumably the Claremont] which together serve to commemorate the passage of time from discovery to the modern day. Soaring cliffs line the far bank and along the near bank; at right, there is a gathering of people (perhaps Native Americans). An information box titled 1609 HUDSON-FULTON CELEBRATION 1909, lists the following 15 events: Sept. 25 Commencement Day N.Y., Sept. 26 Religious Observance Day N.Y., Sept. 27 Reception Day N.Y., Sept. 28 Historical Parade N.Y., Sept. 29 Commemoration Day N.Y., Sept. 30 Military Parade Day N.Y., Oct 1 Naval Parade N.Y., Oct 2 Naval Carnival Parade N.Y., Oct 3. Religious Day Upper Hudson, Oct. 4 Dutchess Co. Day, Oct. 5, Ulster Co. Day, Oct. 6 Green Co. Day, Oct. 7 Columbia Co. Day, Oct. 8 Albany Co. Day, Oct. 9 Rensselaer Co. Day. COPYRIGHT 1909 BY J. KOEHLER, N.Y. [indicated on front only].

Coverage – Geographic: Hudson River, New York and New Jersey

Date Original: 1909?

Publisher: Graduate School of Library and Information Studies – Queens College (CUNY), New York, New York

Language: eng

Source Height: 3.5”

Source Width: 5.5”

Source: Waterways Post Card Collection of Thomas T. Surprenant: Hudson River

Type: Text; Image

Digitization Specifications: Archive masters were scanned at 600 ppi, 24 bit color using a Canon CanoScan 88800F and SilverFast Ai scanning software and saved as uncompressed TIFFs using Photoshop CS\$ and Mac OC 10.5, Derivative access files were saved as medium quality JPEGs, 150 dpi, 1000 pixels on the long dimension. Thumbnails were generated automatically by CONTENT dm.

Date Digital: 2009-09-30

Resource Identifier: h0189ac1.jpg

Format: Image

Format [Medium]: Postcards

File Size: 218972 Bytes

Checksum: 287714794

Height (of Digital Image): 1000

Width (of Digital Image): 637

Color Space: sRGB

Rights Management: Contact Thomas T. Surprenant, ebeltoms@nyc.rr.com

Contributing Organization: Graduate School of Library and Information Studies, Queens College (CUNY), New York, New York

Digitization Team: GSLIS 757, Digital Imaging Fall 2009,: Beth Daniel Lindsay, Meg Donabedian

(GSLIS, 2006-2009, *Waterways of New York V. 2.1*).

Uploading to CONTENTdm

- By mid-semester, the students are introduced to the basics of the CDM Project Client software in a hands-on lab. They upload a sample image, are guided through the creation of a “Waterways of New York” banner to be automatically inserted for future uploads, and—following step-by-step directions—use the Template Creator function to input recurring fields (e.g. Source, Digitization Specifications, Digitization Team), and to enable CDM’s capacity to automatically capture technical metadata such as Resource Identifier, File Size, Format and Checksum fields.
- They then upload another sample image to demonstrate that these functions are working, and to doublecheck (and correct for) possible errors in data entry.

- The following week (assuming appropriate progress on early steps in the workflow), they are ready to upload their first postcard JPEG images (front and back, entered as individual files). Early on in the project we decided against uploading the fronts and backs of cards as compound objects, since this would severely limit the amount of detailed metadata that could be made available.
- Students “cut-and-paste” the doubly reviewed metadata from their Excel files into the appropriate fields in CDM. This approach was especially important when we had limited access to the Project Client interface in the adjacent PC lab, and at times of instability in CDM at the client end.
- Following an additional review by faculty for completeness and accuracy at the desktop, students then upload the files to be approved in the Administrative interface.
- It provides tremendous positive feedback to all to see the fruits of many weeks of work appear live in the evolving database, in real time, during this most momentous class session.
- In subsequent weeks, a portion of each class is devoted to additional project work, reflecting whatever particular step each student is addressing.

Timing /scheduling issues

- Given the pressure of a 15 week semester, and the need to respond quickly to student submissions, the entire quality control process takes much more time than initially projected. Yet, there seems to be no alternative to preventing major errors from creeping into the database. And students regularly report, in mid-year and year-end evaluations, that they find the hands-on nature of metadata creation and feedback to be among the best aspects of the course.
- The time constraints on classes (once a week for 2 hours 35 minutes) make it difficult to teach what is needed and to allow students the time to practice. This is particularly true for commuting students, most of whom hold full-time jobs and may come to school only once a week. This is complicated by

extremely time-consuming commutes (for students living in Brooklyn, three or more hours round trip is common).

- It is not unusual for a few students to still be attempting to scan (or rescan) their postcards during the seventh or eighth week of class (or to have to redo their derivative files due to lack of attention to digital specifications). This puts the faculty—and student partners—into a real bind when it comes their turn to evaluate student work.

How much detail is enough/too much?

- This question has vexed both the faculty and students since the inception of the course. One of the best aspects of the project is the information rich content of the postcards, both back and front. Students have been particularly diligent in finding, and fighting for, new subject terms and descriptive material. This makes the course an ever evolving environment with an endless series of new things to consider and debate.
- Especially dedicated students have taken the time to research historical events depicted in their cards, leading us to add a “References” field in Fall 2008, where they can add more background and even Web citations.
- Student interests, skills, and insights have made this very much a “student-driven” course, in the best sense of the phrase. Every class has made unique contributions to the quality of the project, and all (both faculty and students) learn from one another. The collective contributions of the myriad participants in this project cannot be understated.

Supervisory issues

- As previously noted, the class inevitably encounters difficulties with students who either are forging ahead or are lagging behind others. In any given class the faculty can expect to see a certain level of frustration as students cope with their individual abilities and challenges.

- The evolving documentation, hardware and software problems sometimes create an illusion that the faculty are not well prepared. It doesn't take much to derail what has been planned for any given class. This adds to the timing and scheduling issues discussed above. On the other hand, as more perceptive students note, this is the way projects function in the real world, and provides insights they can share in future job interviews of how they dealt with unforeseen complications.

Role of independent study students in quality control review and overhaul of documentation

- To date, there have been only two Independent Study students interested in assisting with quality control, one for the METRO-funded archival project (Amy Armstrong), and one for the Waterways project (Susan Savage). They both made extraordinary contributions to the projects in terms of documentation and quality control. There is no doubt that, in the near future, more students and/or faculty will have to be brought into the process of oversight and review in order for existing problems and errors to be addressed, but there are substantial challenges in envisioning how this may be possible in the current economic climate.
- The same difficulties exist with the revision of documentation. Incremental changes have resulted in some inconsistencies, and outright errors, in the documentation. At this point having better documentation has a higher priority than quality control review of the existing database. Better documentation will result in better quality control. During the coming year the faculty intend to make major revisions of the documentation to ensure that it is not only up-to-date but more accurate.

Expansion of the database

- We anticipate that, sometime in the near future, the database will have to expand beyond a personal collection. Two preliminary moves in that direction have already been undertaken. The Ellenville Public Library and Museum permitted its Delaware and Hudson (D&H) postcard collection of 70+ cards to be scanned and

inserted into a separate Greenstone database. Similarly, the Erie Canal Museum in Syracuse permitted us to create archival Tiffs of 55 postcards. Students then created derivatives and assigned metadata on these cards, in addition to several they scanned themselves.

- From experience it is, indeed, gratifying to see that a spirit of participation exists within many of the various libraries, museums and historical societies. More needs to be accomplished in this arena and there are plans to work with additional organizations in the coming year.
- Once the collection becomes more mature, enlisting other postcard collectors to contribute to the database is a strong possibility.
- Any future expansion will have to consider moving beyond postcards into other forms and formats. While this is a logical extension, especially with photographs, maps and important historical documentation, the level of difficulty increases. In order to keep students in the process it is likely that the faculty will have to do a great deal of traveling to digitize collections that are all over the state.

Successes and Lessons Learned

The literature of “Best Practices” emphasizes the importance of an “Advisory Council”, and the involvement of key stakeholders, to provide feedback and insight into the successes and challenges of an evolving project (e.g. Chapman, 2000). Our students have exceeded every expectation in this regard, and reinforce the importance of input from a variety of stakeholders in the advisory process. Subject headings and description have expanded from an initial focus on canal-specific characteristics (e.g. tugboats, locks), to many attributes of the environment and daily life in early 20th century New York. Detailed rules and guidelines provide standardized descriptions (developed by the students) of trademarks, logos and postmarks and ensure consistency with AACR2 and other standards.

Perhaps most importantly, students have gained an appreciation for every aspect of the overall planning, processes and implementation pertaining to the creation of at least one type of digital collection (images). This practical experience is embedded in an exposure to key readings in the literature, class discussions, and a detailed case study analysis of an existing collection. In particular, they come to understand the enormous demands of metadata creation and quality control, and the tradeoffs that may lead digital collection managers to minimize this aspect of their projects. The role of students as key project participants has been extremely valuable in providing concrete evidence of their experience as they search for jobs in this difficult economy. It has been equally valuable for potential employers seeking qualified job candidates.

At this writing, the project continues, with two course sections scheduled for Spring 2010 and no end in sight. Our students have gone on to excellent positions in the field and the feedback from intern supervisors is consistently positive. With any luck, our collection will exceed 1,300 images by early 2010.

Challenges and next steps

In addition to many of the future plans outlined above, we are seriously considering migration of the database to an open-source platform, such as Omeka (<http://omeka.org/>), Greenstone (<http://www.greenstone.org/>) or other available options. Running such a digital library collection platform in parallel with CDM would present a valuable learning tool for students, while adding to the complexity of the project.

Possible student interest in working on this initiative as an independent study project, coupled with a soon-to-be-completed search for a new lecturer line at the GSLIS, with specializations in digitization and related technologies, provides some basis for optimism as we move into the next decade. And we remain committed to this labor-of-love that has inspired and empowered so many classes of students.

Yet on the challenges side, for faculty it remains a difficult balancing act. The nature of a project such as this requires continuous updating and editing to ensure consistency and quality in the ongoing development of the project. This process places serious stresses on the competing obligations of supervising faculty for research, publication, service, staying current, and responsibilities for other courses. This is particularly true during a major period of transition within our department. But, some things are very much worth doing, and we hope this example may provide encouragement to others starting out, or expanding their existing initiatives. We look forward to extending the dialogue!

References

- California Digital Library. (2009, September). "3.5 Guidelines for access and thumbnail image files." *CDL Guidelines for Digital Images, Version 2.0*. Retrieved December 31, 2009, from <http://www.cdlib.org/inside/diglib/guidelines/bpgimages/>
- Chapman, S. (2000). Section III: Considerations for project management. In M.K. Sitts (Ed.) *Handbook for digital projects: A management tool for preservation and access*. Retrieved December 30, 2009, from <http://www.nedcc.org/resources/pubs.php>
- Graduate School of Library and Information Studies. Queens College, CUNY. (2005-2009). *Digitization projects*. Retrieved January 3, 2010, from <http://qcpages.qc.edu/GSLIS/digitization.html>
- Graduate School of Library and Information Studies. Queens College, CUNY. (2006-2009). *Waterways of New York V. 2.1*. Retrieved January 3, 2009, from http://cdm128401.cdmhost.com/cdm4/browse.php?CISOROOT=%2Fqcgslis_fo6
- JISC Digital Media. (2008, November 11). *Choosing a file name*. Retrieved September 2, 2009 from <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-file-name/>
- JISC Digital Media (2008, November 14). *Project management for a digitisation project*. Retrieved December 31, 2009, from

<http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/project-management-for-a-digitisation-project/>

Library of Congress. (2007). *Thesaurus for graphic materials I*.

Retrieved December 31, 2009, from

<http://www.loc.gov/rr/print/tgm1/>

North Carolina ECHO Project. (2007). *Project planning*.

Digitization guidelines, 2007 revised edition. Retrieved January 2, 2010, from http://www.ncecho.org/dig/guide_1planning.shtml

Smithsonian Institution. (n.d.). *Greetings from the Smithsonian: A*

postcard history of the Smithsonian Institution. Retrieved

December 31, 2009, from <http://siarchives.si.edu/history/exhibits/postcard/chronology.htm>