Mission: Access to the Library's Unique Digital Resources

The core of DLA's mission is to make the library's unique and rare digital resources available online. This includes hosting born-digital materials as well as hosting scanned materials and providing intermediary services for materials to be scanned, and maintaining a preservation strategy that ensures long-term access. This report highlights these activities for fiscal year 2010. This year DLA concentrated on implementing the library's new digital collection management system, CONTENTdm.

Digital Library and Archives FY 2010 Annual Report
based on the Virginia Tech Strategic Plan: http://www.president.vt.edu/strategic-plan/strategic-plan.html
http://scholar.lib.vt.edu

SCHOLARSHIP DOMAINS
Learning
Increase and support a diverse student body

Students Support DLA's Mission

We are both happy and sad that two of our long-time student workers, Daniel Culpepper and T.J. Rossmeissl, walked at commencement this spring. In anticipation of their departures, early in the year we hired two undergraduates, Kaavya Giridhar and Collin Brittle, who have proven to be outstanding contributors to the work of DLA. DLA benefitted tremendously from Shvetha Soundararajan's second year as the DLA graduate assistant. DLA employed summer student workers: Archana Subramaniam and Bhavesh Jinadra for ejournal markup and Kenneth Black for SAHC. Brian Montanio worked fall semester on the IAWA Database. DLA's students received considerable training, including

- File Permission Management – Kaavya, Archana, Bhavesh, Collin, Ken
- SFTP (Cyberduck) – Kaavya, Archana, Bhavesh, Collin, Ken
- BBEdit – Collin, Kaavya, Ken
- Regex / Grep – TJ, Kaavya, Archana, Bhavesh, Collin
- Command line Linux / Unix functions – Collin, Bhavesh, Archana, Ken
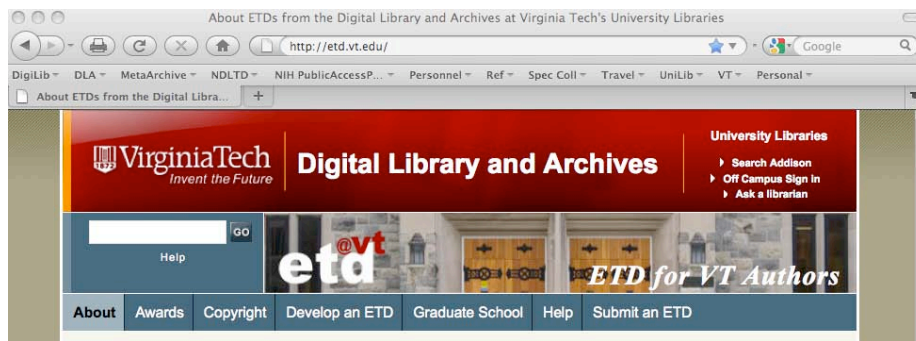- Mac Os 10.5 training – Kaavya, Archana, Bhavesh, Collin, Ken

With this year's focus on the "University Libraries April 16, 2007, Condolence Archives" and CONTENTdm, we staffed DLA with students who not only possessed the skills to learn and use HTML, but who also had beginning programming skills and aspirations. Helping these students to meet their educational and career goals also helps DLA to improve and standardize the metadata, and increase the accessibility and discovery potential of our increasingly automated content.

DLA's student workers contributed well over 2,000 hours to make the library's unique and rare digital resources available online and well preserved.

ETDs: Electronic Theses and Dissertations   http://scholar.lib.vt.edu/theses
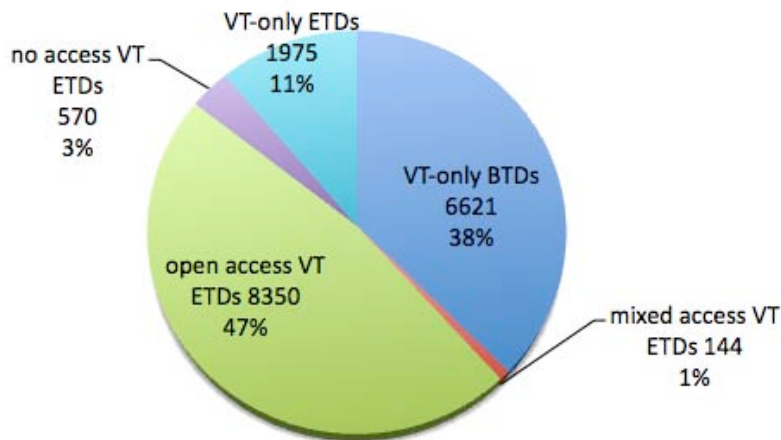
There are many facets to DLA's work with ETDs. Our new systems administrator, Paul Mather, kept the aged software, ETD-db, functioning through crises as well as incremental processing changes to accommodate BTDs. With the support of Library Systems, Sung Hee Park, a graduate student in computer science, began a major reprogramming effort to upgrade the programming language (from PERL to Ruby on Rails), tap into existing online information (e.g., Banner and Plan of Study), improve functionality (e.g., timed removal of access restrictions), and enhance access (e.g., search by any data element—faculty, year, etc.). Shvetha Soundararajan, GA, thoroughly revised etd.vt.edu based on a semester's-worth of meetings with Graduate School personnel and DLA's implementation of the VT Web templates.

BTDs: Scanned Bound Theses and Dissertations

DLA processes ~50 BTDs per week from Circulation/Storage to prepare them for Digital Imaging to scan. DLA also receives ~50 PDF files weekly from DI for uploading to the ETD database. This spring we began OCR'ng BTDs, similar to the *Bugle* process above, but working with PDFs. These files are then uploaded so that computer scripts can process them into the ETD database. DLA also began saving BTD files in the archival format now required of all born-digital ETDs, PDF/A. DLA processed 2,965 BTD, including OCR'ng the last 550.

Availability of VT ETDs as of June 7, 2010

This year we also began initiating requests to BTD authors for permission to remove the VT-only restrictions. Previously we contacted these authors only after we received an external researcher's request. DLA received more than 217 requests for access to VT-only ETDs and BTDs. Following standard procedures, we made 63 more BTDs publicly available this year, but DLA also began releasing ETDs for the Graduate School. As a result, more than 516 ETDs are publicly available that would otherwise still have VT-only access.

IAWA Biographical Database   http://lumiere.lib.vt.edu/iawa_db/

DLA continues to improve the historical record of women in architecture by adding information about their careers to this public resource. This year the IAWA benefited tremendously from the resourcefulness and diligence of volunteer, Adowa Adasi. The database grew by 745 biographies, bringing the total to 2030 entries.

Discovery
Innovative Technologies and Complex Systems

Preservation   http://www.metaarchive.org

Our membership in the MetaArchive Cooperative continues to serve the library well. For example, we received and installed our third server funded by MetaArchive and repurposed the 2004 original. 70 of Virginia Tech's rare and unique digital collections (~215 GB) are distributed among secure dark archives hosted by our partners in the Cooperative. The membership has grown from the original institutions (Auburn, Emory, Florida State, Georgia Tech, Library of Congress, Louisville, VT) with the addition of Boston, Clemson, Folger Shakespeare Library, Hull, Indiana State, North Texas, Penn State, PUC Rio, Rice, and South Carolina.

DLA and, indeed, the MetaArchive technical staff had come to depend on DLA's graduating senior Daniel Culpepper. However, we were able to hire and train Colin Brittle and he is already a very competent plugin developer and data wrangler. Together they developed new procedures for archiving ETDs, which was necessary due to increase in BTD scanning. All of VT's born-digital and scanned theses and dissertations through 2009 are secure in the MetaArchives distributed digital preservation network so that if we experience a catastrophe like last spring's simultaneous double disk failure, we will again be able to recover quickly and thoroughly.

DLA contributed to MetaArchive's NEH grant application. Through "Chronicles in Preservation," the MetaArchive, together with the San Diego Supercomputer Center and the libraries of Penn State University, University of North Texas, Virginia Tech, University of Utah, Georgia Tech, Boston College, and Clemson University, proposes a research and development project to study and document preserving digitized and born-digital newspapers using collaborative and distributed preservation techniques and methodologies. DLA is unique in that it has born-digital newspapers.

Gail McMillan contributed to the MetaArchive Cooperative's book, *Guide to Distributed Digital Preservation;* wrote an article about distributed digital preservation for the Latin American ETD newsletter, *Tesis y Disertaciones Electronicas* (Brazil: PUC-Rio, Dec. 17, 2009); and made a presentation on the same topic to the Canadian ETD & Open Repositories Workshop, Carleton University, Ottawa, May 11.

April 16, 2007, Condolence Archives   http://contentdm.lib.vt.edu/

DLA led the library's initiative to consolidate and provide public access to the condolence materials received by the university following April 16, 2007. We coordinated a two-day training session in July for 17 library faculty and staff and two more from Radford University. All DLA staff actively participated in the 17-member taskforce that met every two weeks through March as well as led the work of several small working groups that tackled everything from installing the CONTENTdm software on the reworked-Solaris server to populating it with digital objects and metadata, and maintaining this new database. We used Scholar to manage documentation, communication, and forums; this now serves as the historical record of this project. Once the taskforce established the policies and guidelines and Ann Lawrence and Jeremiah Puryear developed the Website, the bulk of the work--quality control and uploading--fell to DLA staff, Special Collections' Jenny Mitchell, Technical Service's Jana Doyle, and Art/Architecture's Steve Tatum. 3,905 metadata records were edited to better reflect the digital object displayed without losing the rich information documenting the original artifact. The Condolence Archives has become an attractive and searchable database of over 6,215 (as of June 10) searchable objects.

Ejournals   http://scholar.lib.vt.edu/ejournals/

This year DLA staff and students added 741 articles in 51 ejournal issues, primarily as a PDF file of each article. In addition, we reformatted 253 of these articles in 28 issues as fully accessible HTML.   This provides full accessibility to our files, as required by Section 508 Federal Accessibility Standards. The PDF version offers ease of printing and an exact record of the original article, as it was first published. They also made sure every ejournal web page has server-side footers and headers, which will make future

global updates (e.g., changing Web design) much easier. In addition, they developed new CSS styles, which streamlines the HTML tagging process and ensures more consistent results.

This year was unique in that one of our ejournals, Techné, decided to move to a commercial publisher, making the ejournal available only to paid members of the sponsoring society. We negotiated to provide public access to locally hosted HTML versions of 1995-2008 (vols. 1-12), but the PDF versions of 1-13 are available to VT-only. Virginia Tech will also continue to have access without charge to future issues in appreciation for hosting Techné during its formative years.

VT ImageBase   http://imagebase.lib.vt.edu/

A content-rich information resource, the ImageBase continues to be used by the university community and others. Web-captured images meet the needs of many, but users still requested 241 images, including

- Kornweibel, Theodore. *Railroads in the African American Experience: A Photographic Journey*. Johns Hopkins University Press, 2010.
- Kirby, D.A. *Lab Coats in Hollywood: Scientists Impact on Cinema, Cinema's Impact on Science and Technology*. MIT Press, 2011.
- Exhibitions
  - Manassas Museum
  - Town of Blacksburg's new office building
  - "Pioneering Women Architects" in Richmond
- NASA's Lifetime Achievement Award
- T-Shirts for Black Lung Clinic participants at Washington and Lee's School of Law.

Due largely to our focus on CONTENTdm, we did not add any images to the ImageBase. We were also constrained by old software that needs to be upgraded. Not until May 2010 were we able to update metadata.

> Beginning FY 2011 we would like to remodel our use policy to request a donation rather than charging per image; we believe this would garner the library more than the $503 we received for image use this year. This would also align DLA's policy with Special Collections.

DLA Web Redesign   http://scholar.lib.vt.edu/collections/

Largely through the industrious work of the DLA Graduate Assistant, Shvetha Soundararajan, we made major progress revamping our Web site to conform to the university templates. 64,652 pages reflect the new template. Together with Kimberli Weeks, DLA Technical Director, and other DLA staff, they updated 10,637 ejournal pages, 53,959 news pages, and the majority of the faculty archive, copyright, and collections pages.

One outcome of DLA's early entrance into Web hosting is that over 300,000 of its pages lack, for example, a standardized document type declaration or server-side scripts of standardized information for headers and footers. This year many DLA staff helped to rectify this by adding basic Dublin Core metadata to 53,903 ejournal pages.

Copyright   http://scholar.lib.vt.edu/copyright

The second section of the DLA Web site to move to the new templates was Copyright. Small in comparison, 50 pages, we overhauled this entire section, organizing, checking links, and adding new resources. For example, we developed the Fair Use Analyzer Tool at the request of the Graduate School dean. McMillan developed the questions based on the Fair Use Checklist (http://scholar.lib.vt.edu/copyright/FairUseChecklistVT.pdf) and Shvetha Soundararajan, DLA GA, wrote the scripts to make it a functioning Web tool.

Foundation Strategies
Organizational Development: Recruit, retain, and develop an excellent and diverse workforce

Personnel

DLA had a very stable workforce this year, with only our student workers transitioning as usual. Weeks continued to serve on the Library Staff Association, which elected Anne Lawrence Vice President. Anne also served on the library's Web Team.

Everyone in DLA developed new skills working with CONTENTdm through on-site training, online tutorials and webinars, and informally.

Geri Ingram at CONTENTdm Training. *Photographs by Ann Lawrence, DLA Online Editor*

All DLA staff took part in several professional development opportunities, including two ASERL webinars, In Service Day 2010, and McComas Staff Leadership and Staff Appreciation activities. Many took advantage of various skill development opportunities, including those taught by library personnel (e.g., Travel Center, videoconferencing, SEAS) and others (e.g., IT security, online records management), and attended various vendors' presentations, including Serials Solutions' Summon and Apple's iPad. [Lawrence report 85 hrs, Weeks 63 hrs, and Wills 32 hrs. See also McMillan's FAR 2010.]

DLA from a Systems POV

There are over 5.7 TB of data on DLA servers, which include six real hardware servers, three virtual servers, and ten Web server configurations. There are also three physical hardware servers providing MetaArchive preservation cache services: two production caches and one MetaArchive Test Network cache. There is also an additional physical Sun server that hosts the CONTENTdm local server. The change in servers reflects a deep reorganization that took place in the wake of the May 2009 systemic failure of the DLA servers. The major focus of change has been to remove some of the unnecessary complexity of the original DLA cluster setup and to streamline operations onto fewer machines.

One severe shortcoming of the pre-May 2009 DLA servers setup was the lack of backups and data safety, either through Library Systems or centralized IT. Instead, separate storage in the DLA cluster was repurposed to act as a mirror of the DLA server content. Every night since August 2009, all DLA data is mirrored onto local storage. A mirror is good for quick disaster recovery and failover, but does not allow point-in-time recovery of previous versions of files. So, if a file is accidentally deleted or is edited, there is an extremely short window in which to retrieve the previous version from the mirror. To address this situation, DLA purchased a tape library backup system, which will enable Systems to make regular full and incremental backups of DLA data to LTO-4 tapes. We anticipate having at least a three month window onto daily backups.

Interaction of personnel with the university community and the broader academic communities worldwide benefits the university organization as well as its students.

Collaborations

DLA's mission to make the library's unique resources available is accomplished through collaborations with others. Special Collections and DLA have been operating as separate departments for less than three years, but our cooperation continues to benefit library users. For example, DLA arranged for Digital Imaging (DI) to scan 26 *Bugles*. DLA applied optical character recognition (OCR) to 20 of the high-resolution image files through batch processing using Adobe Acrobat. We posted these for public access at http://spec.lib.vt.edu/archives/bugle/ There are only four volumes remaining to be scanned. We will continue to retrospectively OCR until all 101 Bugles are searchable. DLA also hosts 167 digitized rare books from the Peacock-Harper Culinary History Collection, but these older fonts are much more difficult to OCR so only four are currently searchable. DLA works with Technical Services to ensure the broadest possible access to DLA-hosted resources through Addison and WorldCat. We also work with other units of the library and the university as noted throughout this report.

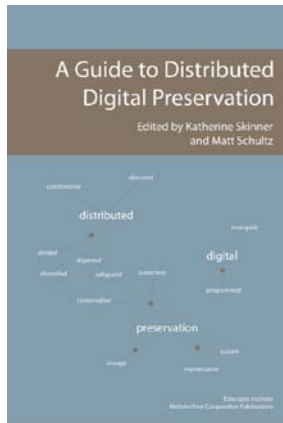## Faculty Archives   http://scholar.lib.vt.edu/faculty_archives/

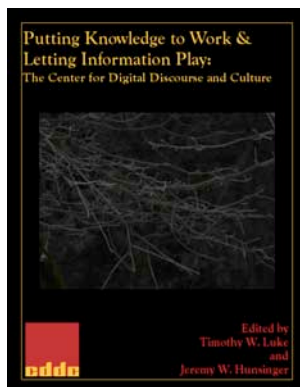Three new collections increased the unique resources available through DLA's Faculty Archives.

- Dr. Elizabeth Fine, Humanities Program Director, is active in the International Colloquium on Communication and received permission from an enlightened German publisher, Röhrig Universitätsverlag, to make the biennial conference proceedings available online. DLA's student web-workers scanned the text-only pages using the office photocopier, OCR'd using Acrobat, and created tables of contents using the HTML skills we taught them. Three issues are now at http://scholar.lib.vt.edu/ejournals/ICC/.

- Dr. Mark Sanders, Technology Education Program Leader, originated Graphic Communications Central (GCC) in 1997 to provide access to educational resources for faculty, students, and trainers. In 2009 DLA captured 1,571 Web pages and packaged GCC into a 61 MB PDF.
- Librarian Philip Young's "The Serials Crisis and Open Access: A White Paper for the Virginia Tech Commission on Research" has worldwide access from the DLA server.

Publications

McMillan, G., Howard, R. (2010) "Content Selection, Preparation, and Management," Ch. 6 of *The Guide to Distributed Digital Preservation,* Katherine Skinner and Matt Schultz, eds. Atlanta, GA: Educopia Institute, 49-71. http://www.metaarchive.org/GDDP

McMillan, G., Fox, E., Srinivasan, V. (2010) "Electronic Theses and Dissertations: Progress, Issues, and Prospects." Ch. 6 of *Putting Knowledge to Work & Letting Information Play: The Center for Digital Discourse and Culture*, Tim Luke, Jeremy Hunsinger, eds. Blacksburg, VA: CDDC.
http://www.cddc.vt.edu/10th-book/

McMillan, G., Ramirez, M. (2010 Jan./Feb.) "FERPA and Student Work: Considerations for Electronic Theses and Dissertations." *DLib Magazine*.
http://www.dlib.org/dlib/january10/ramirez/01ramirez.html

McMillan, G. (2009, Dec. 17) "Adopt a Real ETD Preservation Strategy." *Tesis y Disertaciones Electronicas*. Brazil: PUC-Rio.
http://www.maxwell.lambda.ele.puc-rio.br/acessoConteudo.php?nrseqoco=49349

Presentations

University of Texas at Austin hosted the 13th International Symposium, June 15-18.

https://conferences.tdl.org/utlibraries/etd2010/schedConf/program

McMillan, G. (2010 May). "Collaborative Preservation of ETDs: The MetaArchive Cooperative and LOCKSS" presented at the 1st Canadian ETD & Open Repositories Workshop, Carleton University.

McMillan, G. Halbert, M., Donovan, W. F. "Distributed Digital Preservation Workshop."

McMillan, G., Ramirez, M. "How Privacy Issues Impact Electronic Theses and Dissertations (ETD) Initiatives."

McMillan, G. "Preservation Collaboration: NDLTD & MetaArchive Cooperative." Newcomers Workshop.

McMillan, G., Fox, E. "ETD-db: Today, Tomorrow, and the Future CaR." Newcomers Workshop.

http://scholar.lib.vt.edu/staff/gailmac/MetaA4Canada2010rev.ppt

McMillan, G. (2010 March). "History of the Peacock-Harper Culinary History Collection," presented at the 10th Anniversary of the Peacock-Harper Culinary History Collection, Roanoke, VA.
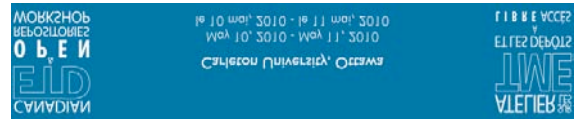http://scholar.lib.vt.edu/presentations/PHCHC5minutehistory.ppt
http://scholar.lib.vt.edu/CulinaryHistory/PHCHC10th/PHCC10th.mov