

**Digital Archiving Review Taskforce
Final Report
November 11, 2004**

The Digital Archiving Review Taskforce was charged with assessing digital archiving needs for objects accessed through the Digital Library at Dartmouth and stored within the library and campus systems. The charge issued by the Digital Library Management Group specified these key deliverables:

Inventory of digital objects and their current storage and backup methods including any existing guidelines or standards being used to ensure permanence.

Recommend methods for creating a program to ensure permanence for digital objects including suggested structure and membership. Recommendation should cover already created digital objects as well as provide clear guidelines for anyone creating digital objects for permanent retention.

A survey was designed and placed on the web during late March and early April [see Appendix 1]. Over 300 items were reported and a spreadsheet was created by the taskforce listing all items reported on the survey [see Appendix 2]. The taskforce met from late April until early June to review the gathered information and assess the archival state of digital information.

The items reported on the survey fall into several large categories, the largest group consisting of items served via Dartmouth College Information System (DCIS). This includes Special Collections finding aids and databases, electronic texts owned by Computing Services, electronic texts owned by the library, and databases owned by academic departments. Also reported were four digital publications; a number of items held only on CD; Subject Guides; and a couple library-produced databases. Back up of digital items was handled within the department that created the digital objects. There did not appear to be any guidelines for permanence.

The resources available through the DCIS system are being addressed by the DCIS Database Migration Plan that is being managed by Jennifer Merrill and John James [See Appendix 3]. Of primary focus are the databases utilizing the BRS database management system. Some of these databases are being replaced by commercial solutions or are simply being eliminated, as they are no of use to our patrons. Some resources will migrate to more modern solutions. Those will remain within the purview of this group and their solutions will comply with the policies and guidelines set forth by this or subsequent groups. Because the Digital Libraries Technologies Group is working directly with data owners and because a plan and timeline are in place the taskforce did not address the preservation of these items.

Recommendations:

1. All digital objects accessed through the Digital Library should be served from a production server.

The taskforce agreed on the need to have digital items stored and served from a computer that is available 24 hours and is regularly backed up by the Computing Services [see Appendix 4]. The preferred method is for the library to store data in a digital repository. This repository would be a centrally managed, production server owned by the library and dedicated to digital content. For that reason the taskforce recommends that all digital objects accessed through the digital library should be served from Webster or another appropriate server until a repository is created. Digital content should not be stored on an experimental test bed nor should it be kept on departmental computers.

Some Subject Guides are served from a librarian's personal web space. This should be stopped in favor of serving Subject Guides from the Research and Instructional Service web space [~bakerref/], which is served from Webster or another appropriate server.

2. Policy should be written regarding the use of college server space by library staff for professional organizations.

There were two items reported on the survey that were web pages hosted by a librarian for a professional organization. The taskforce recommends that a policy be written regarding the use of college server space by library staff for professional organizations and that if the use of college resources is endorsed that a campus supplied central web server be used.

3. Policy should be written regarding the disposition of digital projects when the creator has left campus.

Another area of concern was the disposition of digital resources created by library staff after the staff member has left the community. The "Sor Juana" files fall into this category and are addressed in the DCIS Database Migration Plan. However, it is recommended that a policy be written regarding the disposition of digital projects when the creator has left campus so that anyone planning to leave understands their responsibilities in regard to the project they have created.

4. Standards must be easy to locate and training should be provided for staff to understand their use.

It was observed by members of the taskforce that many of the objects identified in the survey are tools used by departments rather than objects accessible via the Dartmouth Digital Library. As tools they may not have permanent archival value and adherence to standards of permanence is not a critical issue. The tools should be examined to ensure that they migrate forward to modern versions or be removed if they are no longer needed.

Another concern of the taskforce is that digital objects are being created for permanence but the standard is not being followed. Standards for the creation of digital objects were created in January 2002 and are posted on the Digital Library Management Group web

page¹ but have not been faithfully followed. We recommend that staff be provided with clear guidelines as to how standards should be utilized. Workshops should also be offered to familiarize staff with standards and how to implement them.

5. Appraisal and selection policies must be created to identify the digital objects that should be kept in the digital repository.

The key factor in saving digital content for long-term use is deciding which content will be saved. This taskforce was charged with inventorying what had been created and how it was saved. Server space and the long-term maintenance of the objects being kept are expensive long-term investments. A program to ensure permanence should include appraisal and selection. We do not recommend that everything identified in the survey be automatically added to the digital repository. The Digital Library Group working with the Collection Management and Planning Group should appoint a taskforce to develop selection criteria for the digital repository and then determine which items from the inventory should be selected for permanent storage.

6. Develop policies addressing infrastructure, standards, metadata, selection, and migration.

The taskforce does not have a specific recommendation for creating a program to ensure permanence for digital objects. Large issues must be addressed by the Digital Library Group before digital archiving can be established in a coherent and rational manner. Chief among these are infrastructure, standards, metadata, selection for retention, and supporting policies. The library has a relatively small number of digital objects and can still maintain hopes to create a system that will ensure long-term preservation.

Through the work of selection the repository can grow. The technical aspects of creating digital content can be disseminated through the Digital Library Group website and workshops. A working group may be formed as needed to review and update standards as technology develops. Through selection, standards, and a repository the library will be able to create a secure space to store the digital object important to the research needs of the Dartmouth College Community.

~~~

**Digital Archiving Review Taskforce**

Peter Carini

John Cocklin

Jennifer Merrill

Barb Sagraves, chair

November 11, 2004

---

<sup>1</sup> URL for Standards: <http://diglib.dartmouth.edu/libserv/digpub.shtml#standards>

## **APPENDICES**

### **Appendix 1**

#### **Web Survey** (Text from web survey)

#### Digital Objects Survey

The preservation of digital content is an important part of the library's mission. In order to identify preservation needs and to plan for future digital preservation the Digital Archiving Review Taskforce (DART) is conducting a survey to identify digital content which the library has an on-going responsibility for preservation. The goals of the survey are to identify digital objects and the current method of archiving being employed. For the purposes of this survey a digital object is defined as anything that contains digital content such as text, images, video, or metadata. They may be stored on a library or campus server, a desktop hard drive, and may be born digital at Dartmouth College Library or commercially produced and selected for acquisition and served to the public by the Library.

Some examples of digital content that should be reported include:

- Scanned manuscripts, maps, text, etc.
- Any object scanned for duplication
- Subject guides
- Local databases
- Electronic journals published by the library
- Analog materials reformatted to digital
- Commercial product purchased or obtained by the library and served to the public from library servers.

Please take a few minutes to respond to this survey. We would like:

1. The name of the digital object or collection.
2. Brief description of the digital object or collection. (optional)
3. URL (if available) or location of computer from which you access the digital object.
4. Enter a brief description of the archive (backup) method employed.
5. The type of file format (i.e. JPG, TXT, WAV)
6. Is it owned by the Library in a non-digital format (y/n)
7. Who should be contacted for more information?

Please list as much information as possible. If you aren't sure if a digital object should be submitted please contact a member of the survey group. We expect overlap and would rather a collection is over reported than left out. The deadline for reporting is April 16, 2004.

Thank you for your cooperation.

Digital Archiving Review Taskforce: Peter Carini, John Cocklin, Jennifer Merrill, Barb Sagraves (chair)

## Appendix 2

**Digital Objects List** – spreadsheet is separate enclosure “DigObjListRESIVED.xls”

## Appendix 3

### DCIS Database Migration plan

1/04 Notify Tuck Career Services they have until 6/1/04 to migrate or remove their database **DONE they are ready for us to remove on 7/1/04 – Removed 7/12/04**

1/04 Meet with Jean Kinney to discuss migration of CORK database – John, Jennifer **DONE expect answer by end of February, and gave her until summer to remove the database. 9/14/04 Gave Jean deadline; database will be removed on Dec 20, 2004.**

3/04 Ask Miguel whether Sor Juana files are wanted for scholarship at Dartmouth. If not, write to Luis and tell him we are changing systems and can no longer host the databases – John

3/04 Refer Material Safety Data Sheets to Barbara for replacement suggestion – John **DONE early summer 2004. 9/14/04 John will touch base with Barbara.**

3/04 Meet with RIS to discuss commercial replacement options for: the 3 Bibles, 2 Frost [might have a Rauner interest], New Hampshire Newspapers, 3 Shakespeare, all of "Malcolm's files" (which are in the library catalog). United States Constitution – move to web version? <<http://www.house.gov/Constitution/Constitution.html>> Goal is to have a plan for what to do by end of spring term (even if we can't fund it all right away.) **9/14/04 Removing Frost from the list as these are Lathem's and we have an interest. Possible XML solution in the future? John will take rest of list to RIS soon.**

3/04 Develop strategy for migrating to LION as a commercial replacement for the 11 Chadwyck-Healey databases (African-American Poetry, African-American Poetry Bibliography, American Poetry, American Poetry Bibliography, English Poetry, English Poetry Bibliography and the 5 Goethe databases) in time for FY05 budget process. **DONE – LION purchased 6/04. 9/04 working on local database demise.**

3/04 Develop plan for Photo Records with Rauner, Cataloging, and DLTG. Plan needs to serve needs of Archives for the full collection as well as any public access to the collection.

3/04 Develop plan for G Major, Playbills, Thespis [to migrate to Finding Aids system, some other system, or go away] – Cataloging, Rauner, Jennifer

3/04 Develop a local solution for Environmental Studies Catalog, Ethics Institute Bibliography, and 2 Native American Studies catalogs, Women's Resource Center Catalog – DLTG+John

3/04 Talk to Hood about their system and are they thinking about it's lifetime, future needs, etc – John, Jennifer **DONE, short term solution is to move from BRS to DII, long term solution dependent on the HOOD**

3/04 Academic Computing explores migration of Video Encyclopedia database.

9/04 Dante migrates to Oracle [or some other solution of Computing's choice] – Steve Campbell project begins 1/04, estimated 6 months of work, goal to be off by fall term

DCIS Databases  
as of 9/14/2004

### **Library Databases**

Bible - New International Version - PAT

Bible - Revised Standard Version - PAT

Bible - Vulgate Version - PAT

Chadwyck-Healey African American Poetry - PAT

Chadwyck-Healey African American Poetry Bibliography - PAT

Chadwyck-Healey American Poetry - PAT

Chadwyck-Healey American Poetry Bibliography - PAT

Chadwyck-Healey English Poetry - PAT

Chadwyck-Healey English Poetry Bibliography - PAT

Chadwyck-Healey Goethe Conversations - PAT

Chadwyck-Healey Goethe Literary Works - PAT

Chadwyck-Healey Goethe Diaries- PAT

Chadwyck-Healey Goethe Letters - PAT

Chadwyck-Healey Goethe Scientific Works - PAT

Frost, R. Poetry – PAT – The Edward Connery Lathem edition

Frost, R. Poetry Masques – PAT – The Edward Connery Lathem edition

Library Staff Directory – **BRS In transition off BRS 9/2004**

New Hampshire Newspapers - BRS

Photographic Records Catalog – BRS

Shakespeare Plays - BRS

Shakespeare Sonnets - BRS

Sor Juana Inez de la Cruz Plays - PAT

Sor Juana Inez de la Cruz Poetry - PAT

Sor Juana Inez de la Cruz Prose - PAT

Special Sheet Music (G Major) - BRS

Theatre Clippings Collection (Thespis) - BRS

Theatre Program Collection (Playbills) - BRS

### **Partner Project Databases**

Career and Employment Services Library – **BRS last update 3/31/2000**

Dartmouth Dante Project - BRS  
Environmental Studies Program Library Catalog – BRS **last updated 7/24/2002**  
Ethics Institute Bibliography – BRS **last updated 4/19/2002**  
Hood Museum Fine Art and Anthro Collections - dii  
Hood Curator Databases (Anthro & Fine Arts) – dii  
Hood Museum Nazi-era objects - dii  
Material Safety Data Sheets - BRS  
Native American Studies Catalog – BRS **last updated 5/22/03**  
Native American Studies Film Catalog – BRS **last updated 1/23/01**  
Project CORK Catalog – BRS  
Video Encyclopedia of the 20th Century – BRS – **last updated with new video 10/30/2001**  
Women’s Resource Center Catalog – BRS – **last updated 12/2003**

### **Administrative or Computing Databases**

|                                         |                       |
|-----------------------------------------|-----------------------|
| Aquinas, T - PAT                        | Spinoza, B - PAT      |
| Aristotle - PAT                         | Thoreau, H - PAT      |
| Bentham, J - PAT                        | Wittgenstein, L - PAT |
| Burke, E - PAT                          |                       |
| Cather, W - PAT                         |                       |
| Cervantes, M - PAT                      |                       |
| Clemens, S - PAT                        |                       |
| Descartes, R - PAT                      |                       |
| Emerson, R - PAT                        |                       |
| Faculty Handbook - PAT                  |                       |
| Faulkner, W - PAT                       |                       |
| Federalist Papers - PAT                 |                       |
| Hobbes, T - PAT                         |                       |
| Kant, I - PAT                           |                       |
| Kierkegaard, S - PAT                    |                       |
| Leibniz, G - PAT                        |                       |
| London, J - PAT                         |                       |
| Machiavelli, N - PAT                    |                       |
| Melville, H - PAT                       |                       |
| Mill, J - PAT                           |                       |
| Nietzsche, F - PAT                      |                       |
| Paine, T - PAT                          |                       |
| Philosophy 12 Course Reader - PAT       |                       |
| Philosophy Compendium - PAT             |                       |
| Plato - Hamilton and Cairns - PAT       |                       |
| Plato - Jowett - PAT                    |                       |
| Ricardo, D - PAT                        |                       |
| Shakespeare (individual play files) PAT |                       |
| Sidgwick, H - PAT                       |                       |
| Smith, A - PAT                          |                       |

Dartmouth College Administrative Guide - Oracle  
United States Constitution – PAT [**this db available at**  
**<http://www.house.gov/Constitution/Constitution.html>**]

### **Possible database replacements**

Material Safety Data Sheets - <http://hazard.com/msds>

### **Removed databases**

Tuck Career Services Library – BRS **Removed 7/13/04**  
Serials Worklist file (SWRK) – BRS **REMOVED 7/13/04**  
PTSD Resource Center – BRS **Removed 12/2003**

### **Appendix 4**

#### **Backup procedures for servers residing in the Machine Room that includes Library central servers.**

Incremental backups are done every day, except the day the Full backup is run. An Incremental backup writes out every file that was changed since the last Full backup. Full backups are done once a week. Once a month the Full backup for that day is pulled and kept for a year from the date it was made.

Backups are written on a Tape Library in another building, and then returned to the machine room for storage after two months. Tapes do not go off campus.