Digital Library Technologies (DLT)
Annual Report, July 2004 - June 2005

CONTENTS

This report describes the activities of Digital Library Technologies (DLT) during the past year. Significant accomplishments include a large increase in the number and quality of information resources available through LIAS, addition of linking tools to tie citations to available full text articles, and significant additions to the emerging Digital Library Infrastructure (OAI, SFX, MetaLib, and DLXS). The past year also saw continued effort on DLT's part to enhance and customize the delivered Sirsi Unicorn library management software to meet the Libraries needs.

Produced by Digital Library Technologies, The Pennsylvania State University, E3 Paterno Library, University Park, PA 16802

http://www.dlt.its.psu.edu/Report.pdf

Information Access

New Databases/Resources. Seventy-two new databases and other electronic resources in a wide variety of subject areas were made available via LIAS during 2004/2005 bringing the total number of e-resources accessible via LIAS to 500+.

A sampling of new electronic resources added in 2004/2005 includes: Gutenberg-e, International Index to the Performing Arts, Opposing Viewpoints Resource Center, Science of Synthesis (Chemistry), Gale Virtual Reference Library, Twentieth-Century Drama, The Nation Digital Archive (1865-), Sociometrics Social Science Electronic Data Library (SSEDL), HAPI (Hispanic American Periodicals Index), International Index to Black Periodicals Full Text, Odum Institute's Public Opinion Poll Question Database, Society of Petroleum Engineers eLibrary, SodaPop, American Geophysical Union e-journals, Early American newspapers (1690-1876), and AnthroSource.

Addition of Large Bibliographic Files to The CAT. For the last few years, *The CAT* has been enhanced by the purchase of collections of bibliographic cataloging data from external sources. Some of these collections provide for the first time full access to materials in the University Libraries for which access was previously minimal or non-existent. For example, a large microfilm collection may have thousands of unique articles or documents that were previously identified in *The CAT* by only one bibliographic record – the title of the collection itself. By purchasing the records that identify each article or document, the usability of the microfilm collection is greatly enhanced.

Did You Know?

26,469 bibliographic records were loaded into The CAT in 2004/2005:

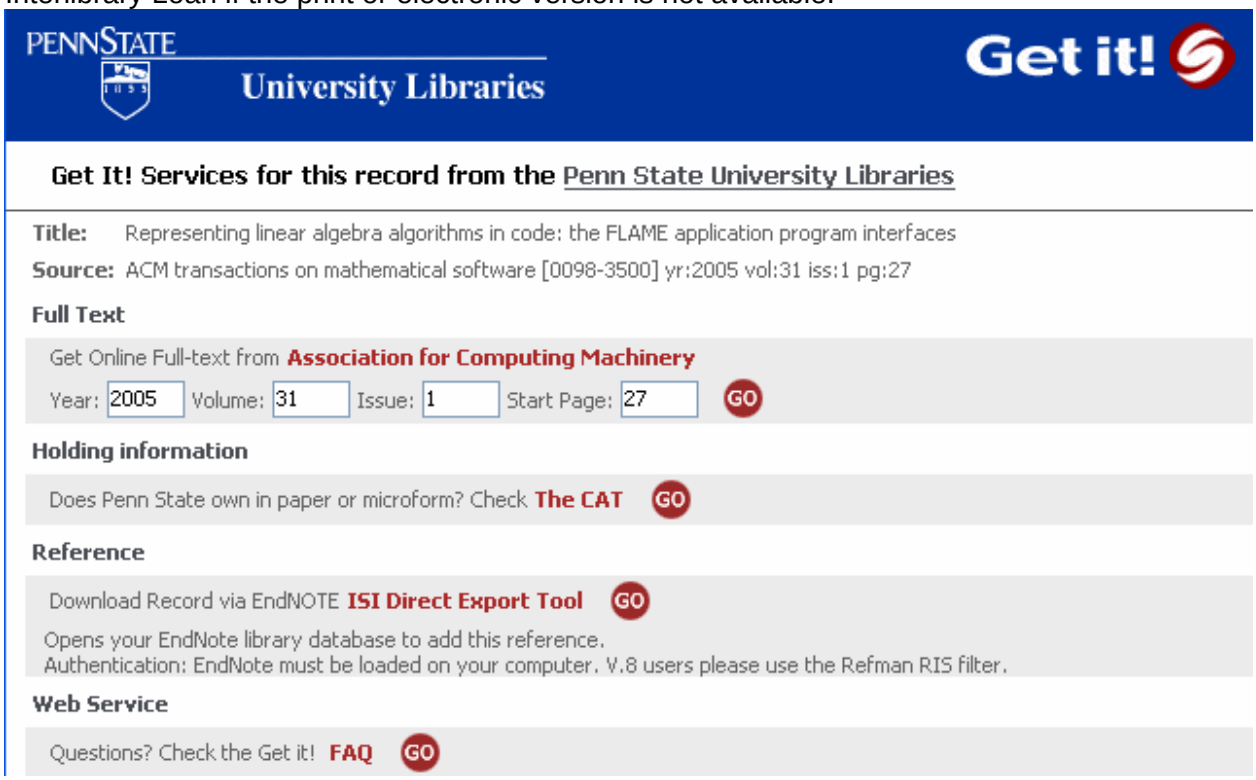
CRCnetBase	212
Papers of NAACP	4,326
American Periodical Series	1,383
ICPSR	5,811
Major Studies and Issue Briefs Microfilm	9,576
Earthscape (Columbia University)	41
CIS Congressional Committee Hearings, Part 1	403
Digitized Maps from Library of Congress	4,717

This means of enriching *The CAT* with pre-cataloged collections will continue. In FY 2004/2005, more than 26,000 titles were added in this way.

Linking Software. In conjunction with the Library, DLT implemented a new service, GetIt. 

Get it! is a new library service that: provides direct links from a citation in a database to online full-text of an article (when available); provides a link to search in The CAT to see if Penn State owns other formats of the publication; and allows users to request an item via

Interlibrary Loan if the print or electronic version is not available.



OAI Harvesting: The CIC Libraries have a Mellon grant to create and implement an OAI-PMH metadata harvesting service to aggregate metadata describing information resources held by participating libraries. The purpose of the grant is to make this metadata aggregation available to end-users (students, faculty, and the general public), both within and outside of the CIC, using appropriate, state-of-the-art search and discovery tools and browsing and navigation interfaces. DLT has provided metadata through an OAI-PMH interface for the Digital Image databases in ContentDM and for the Penn State Electronic Theses and Dissertations database to the CIC-wide metadata harvesting service.

Research and Development

Federated Searching. DLT is working with the Library to implement a new service, MultiSearch. The service is implemented with MetaLib, a federated search engine from ExLibris. This software allows users to cross-search databases and view merged result sets. A soft launch of the MultiSearch is anticipated in the late fall of 2005.

Shibboleth. DLT continues with Emerging Technologies to investigate using Shibboleth to access Electronic Resources. It is anticipated that a new pilot project will begin this fall.

Zebra. DLT loaded three databases on the Solaris platform using the open source product Zebra. This product will provide Z39.50 searching of these databases and will replace a similar service offered on the VMS platform, which is being phased out. The first phase of testing is complete and it is anticipated that two of the three resources, The Daily Collegian and Centre Daily Times will be in production by September.

DSPACE. DLT has installed DSPACE and created a digital repository for the staff in DLT and the Library to investigate more fully the requirements of supporting a digital institutional repository.

DLXS. DLT installed a test and production environment for DLXS. Digital Library Extension Server (DLXS) is a foundation/framework for digital libraries. The Library and DLT are working together migrate the Special Collections Finding Aids into DLXS. These will be made available as part of the Digital Library Infrastructure in the coming year.

Greenstone. Greenstone is a suite of software for building and distributing digital library collections. DLT installed Greenstone on the Solaris platform. The Library is currently using Greenstone to support a few collections on a test basis. These include several joint projects with the Penn State Press.

LionShare. DLT exported two image collections from ContentDM with their associated metadata for access via LionShare. The collections will been installed on a LionShare peer server for access by the LionShare community at Penn State. One collection is an open collection available to all users. The second collection will be accessible only to the Penn State community.

Software Investigations. DLT is investigating several open source products to support the Library's needs.

- Fedora: a digital institutional repository
- WebAccess: authentication to web resources using CoSign
- Zope: Content Management System

SSN Conversion Project. DLT successfully completed the SSN conversion in December.

Other Activities. A variety of other software enhancements not directly related to LIAS were completed during the year. For example, the joint Libraries/DLT Helpdesk was migrated from Tru64 Unix to Solaris UNIX. TechSmart, the Library's class scheduling tool was migrated to Solaris Unix.

Maintenance and Support

Sirsi Unicorn Management and Maintenance. The past year a significant amount of time was dedicated to the Unicorn 2003 upgrade. This release added a number of enhancements and new functionality to the Libraries Integrated Library System, mostly supporting staff operations but also providing for enhanced indexing in *The CAT*. DLT also worked with SIRSI to install, test and migrate Unicorn to Oracle 9i from the older Oracle 8i platform.

DLT and the Libraries implemented a new Web interface for *The CAT* based on the Sirsi iLink product. The new service includes several new features to enhance the searching experience for students and faculty, including:

- a Quick Search for a google-like interface to the library catalog

- an alert service for user-designated additions to the catalog. A monthly email informs the user of new items added to the collection based on an author's name or subject that the user designates.
- two new features to help refine or broaden catalog search requests. The "Limit These Results To" links help tighten down the search by limiting to a particular subject or area. The "Try These Too…" links provide subjects related to the search that can help find more resources.

Infrastructure

New Hardware and Upgrades. DLT performed the following during FY2004/2005:

- Added 1 additional enterprise level SUN server to enhance library applications as well as an additional server to support Cold Fusion applications utilizing Oracle.
    - SUN V6900 with 16 dual core Sparc processors, 64GB of Memory and SAN attached storage.
    - SUN V240 with 2 processors and 8GB of memory.
- Added H/P Storageworks Enterprise Virtual Array (EVA) that provides an additional 42 Terabytes of disk storage.
- Added 3 additional HP Intel based servers running Windows 2003
    - 2 servers utilized for SIRSI Directors Station software
    - 1 server utilized for MTSS Media site Live
        - http://live.libraries.psu.edu
- Upgraded all core network services to GB Ethernet in all University Park Libraries. This upgrade will allow us to decommission all private fiber to branch libraries at University Park once port authentication is solved.
- Added Foundry Networks wireless hardware for the new Physical and Mathematical Sciences Library as well as enhancing coverage in Pattee and Paterno Libraries.
- Added additional Tape storage to our newest tape robot.
- Upgraded 235 LIAS public PC's at all campus locations.

Help Desk. 4424 trouble calls were reported to the joint DLT and University Libraries Help Desk from July 2004 through June 2005. This represents over a 6% increase from the previous year. All University Libraries hardware/software related trouble calls are handled by DLT.

New Position. DLT has added a new Systems Administrator that has primary responsibility for Solaris system administration. This is in conjunction with our transition from VMS and H/P Tru-64 operating systems to SUN Solaris. Backup for individuals has also been enhanced with this new position.

Future Initiatives

Disaster Recovery Planning. DLT will be working closely with Emerging Technologies to test disaster recovery scenarios for off site replication of data and server systems to ensure minimal if any downtime if an unrecoverable failure of systems in the DLT computer room occur. In addition we will be working on a comprehensive plan integrating with all of ITS.

UNIX Server Hardware. DLT will add an additional enterprise level SUN Solaris server to enable greater redundancy and fail over for the current library systems.

Windows Server Hardware. DLT will upgrade several Intel based servers and explore alternatives to these systems. We will be exploring rack and blade server technologies to replace current single cabinet floor standing models. We have received Dell supplied blade server hardware that we will be implementing as our domain for all LIAS Public PC's.

Software Monitoring and Recovery Tools. DLT will be testing software from Altiris Corporation in attempts to improve efficiency in updating, inventory, and recovery of desktop systems.

Diversified Operating Systems. DLT will continue to investigate other operating system platforms in an attempt to lessen the dependency on the Windows Server and Workstation platform. Platform examples are: Novell Netware, SUN Rays, Macintosh and Solaris X86.

Additional Storage for Repositories. DLT will investigate the ever increasing demand for storage in the library environment. We will continue to leverage new technologies in attempts to provide reliable long term storage solutions.

EMAIL and SPAM Software. DLT will continue to upgrade email and web-mail capabilities for the Libraries as well as spam filtering. Current plans include moving EMAIL and SPAM filtering software from our current VMS platform to Solaris to be able to decommission our current dependency on VMS thus reducing maintenance costs.

Single legacy VMS Machine. DLT will purchase a much smaller scale most likely single processor Alpha server to run our only supported VMS application for MTSS. This will save ongoing maintenance expense for hardware and software support.

Day to Day Operations

Ongoing Support. The following tasks are not considered new but support for the technologies that have been implemented.

- Server systems monitoring and updating as needed
    - 2 Enterprise level H/P Tru-64/VMS Servers
        - 1 16 processor
        - 1 8 processor
    - 1 Development Tru-64 Server
    - 1 Enterprise level SUN Solaris Server
        - 16 processor
    - 25 Intel based Windows Servers
        - Several multi-processor
    - 2 Mid tier SUN Solaris servers
        - 8 processors each
    - 1 Development SUN Solaris Server
    - 2 Linux Servers
    - 2 H/P Enterprise Virtual Arrays (EVA's) 54 Terabytes
        - 1 12 Terabyte Fiber Channel drives

        - 1 42 Terabyte Fiber ATA drives
- Helpdesk Support
    - Maintain all DLT and Libraries desktop hardware (faculty, staff and public access) totaling approximately 1600 workstations, including life-cycle activities and salvage.
- Networking for all UP Libraries
    - Maintain All hardware
        - 800 open port (public access) connections
        - Public Wireless access at the Pattee and Paterno Library (all floors), Physical and Mathematical Sciences Library and Pollock Library.
        - All Faculty, Staff and Public (student) wired ports
- Backup of all systems to tape
    - 2 tape robots
- Assist development staff in the installation of new programming environments and systems as needed.

Staff Development

During the past year, DLT staff participated in a total of 3,135 hours of professional development activities. The activities ranged from self-paced reading, completion of web tutorials, and hands-on exploration of new software to more formal activities such as ITS' Teaching and Learning with Technology seminars, HRDC and Mastering Supervision classes, and even some week long workshops at locations other than PSU. Topics that DLT staff studied during the year included: Oracle, Java, XML, SQL, Perl, SSH, Zebra, SFX, MetaLib, Solaris, and more. DLT staff also attended and participated in a number of University Libraries Technology Forums and presentations on new software services.

DLT Professional Development Activity

- 3,135 total hours
- 256 Conferences/Meetings
- HRDC Training - 7 classes
- Web-based Training - 17 courses
- Courses: Perl, Zope, Java, SFX, MetaLib, MS Server 2003, Solaris, Oracle, and others

Appendix A. LIAS Statistics at a Glance

Did You Know That ...	
• 72 new databases and resources were made available through LIAS during 2004/2005.	
• 3.5 million electronic resource "sessions" were logged by users in 2004/2005, a 41% increase in the last 3 years.	
• The CAT accounted for 40% of LIAS e-resource usage during 2004/2005. Usage of The CAT has increased almost 40% in the last 3 years.	
• 4,424 Helpdesk tickets were reported and resolved, a 6% increase over the previous year.	
• 419,139 bibliographic records were added to *The CAT* in the last two years, a 17.4% increase	
2004/2005 in Numbers...	
Databases and Other E-Resources	
Number Available	500+ (16% increase)
Added in 2004/2005	72
Total Number of Sessions	3,532,816 (8% increase)
The CAT	
Bibliographic Records in Database	2,828,732 (5.5% increase)
Bibliographic Loaded in 2004/2005	26,469
Total Number of Sessions	1,404,771 (1.7% increase)
Circulation	
Items in Database	5,190,680
Users in Database	149,591
Charge Transactions (Charges and Renews)	640,036
Hardware	
LIAS Express Workstations in the Libraries	80
Authenticated Public Workstations in the Libraries	670
Staff Workstations in the University Libraries	800
Disk Storage	54 TB (145% increase)
Servers	25 Windows, 9 Unix, 1 VMS

Appendix B. LIAS E-Resource Usage Statistics

LIAS Database and Other E-Resource Usage Statistics, 2001 - 2004

	2001*	2002	2003	2004
Totals				
Total No. Of Sessions Across All Resources	2,542,038	3,055,648	3,272,611	3,576,436
The CAT				
CAT - PSU Libraries Web Catalog	1,103,812	1,350,682	1,381,566	1,404,771
Top Twenty (based on 2003 usage)				
(Statistics for all resources are available at http://www.lias.psu.edu/stats/contents.htm)				
CAT - PSU Libraries Web Catalog	1,016,243	1,350,682	1,381,556	1,404,771
ProQuest Direct (newspapers, magazines, business)	214,786	285,101	320,069	326,257
Electronic Reserves	36,310	14,162	49,151	99,323
PsycINFO (psychology)	52,859	70,901	78,479	86,364
LEXIS-NEXIS Academic Universe	45,081	90,046	86,097	76,872
Web of Science	46,328	70,246	69,968	70,227
JSTOR (full text journal articles)	42,724	56,830	61,413	62,746
ProQuest Psychology Journals	--	--	48,650	58,340
Sociological Abstracts	18,430	25,826	36,317	45,174
Elsevier ScienceDirect Web Editions (fulltext journal articles)	32,836	48,234	42,825	43,812
WorldCat	19,582	38,401	50,243	41,815
ABI/INFORM (business)	24,473	33,220	38,277	38,577
Hoovers Online (Company Information)	3,755	12,488	20,625	26,698
Compendex (Engineering)	11,125	16,047	20,936	26,123
IEEE Electronic Library (IEEEXplore)	26,682	27,902	25,285	24,649
CQ Researcher (Contemporary Isssues)	--	12,603	23,979	23,954
Dow Jones Interactive (now Factiva)	22,876	29,621	26,793	23,426
Blackwell Science/Blackwell Publishing Journals	--	18,817	20,690	22,762
Education Abstracts Full Text [Wilson]	--	13,623	22,785	21,363
Elsevier Science Direct (full text journal articles)	--	16,705	21,135	20,498

Note:
**Data for October - December 2001 were not available. Annual statistics for 2001 were calculated using the data from the first nine months.*