

University Libraries Library Systems Office



Library Technology

Presidential Task Force Briefing

December 8, 2004

Wally Grotophorst
Associate University Librarian
Library Systems Office
University Libraries
George Mason University

The Library Systems Office (LSO) is responsible for the development, installation, maintenance and enhancement of all information technology tools and systems within University Libraries.

The ITU is responsible for network infrastructure.



- Sun SunFire V880 (8 CPU) Voyager
- Sun E3000 (4 CPU) Voyager
- Sun E250 (2 CPU) Library Web / MySQL
- Sun E250 (2 CPU) SunRay Thin Client
- Sun E250 (2 CPU) MasonLink+
- Sun E240 (2 CPU) SunRay Thin Client
- Mac XServe G5 (2 CPU) MARS
- Sun Sparc 20 Intranet, mail relay
- Sun Ultra 2 (2 CPU) EZProxy
- Locally built server (2 AMD Opteron) Web Services
- 350+ Public and Staff workstations in 4 libraries

Software Environment

- Solaris 8
- Linux (SuSE, RedHat)
- BSD Unix
- Mac OS X Server
- Mac OS X (Panther)
- Windows 2000 Server
- Windows XP, Windows 2000

fopen("ppmenc.doc",

Oracle, MySQL, PostgreSQL

exit(2);

- Apache / JBoss / Tomcat
- PHP, Perl, Samba
 - Delphi * Epm = Lopell | Ppl. | | | Error: |

Milestones

- Online circulation system installed 1981
- First PC arrives 1982
- Card catalog replaced 1985
- NOTIS installed 1989
- CD-ROM network installed 1992
- First UNIX server installed 1993
- University website (www.gmu.edu) 1994



Milestones

- Voyager installed 1997
- E-Reserves system developed (OSCR) 1998
- SunRay ThinClient installation 2000
- MyLibrary@Mason 2001
- E-Journal Finder 2002
- Participant in LOCKSS project 2002
- OpenURL Link Resolver (MasonLink+) 2003
- MARS (Digital Repository) 2004

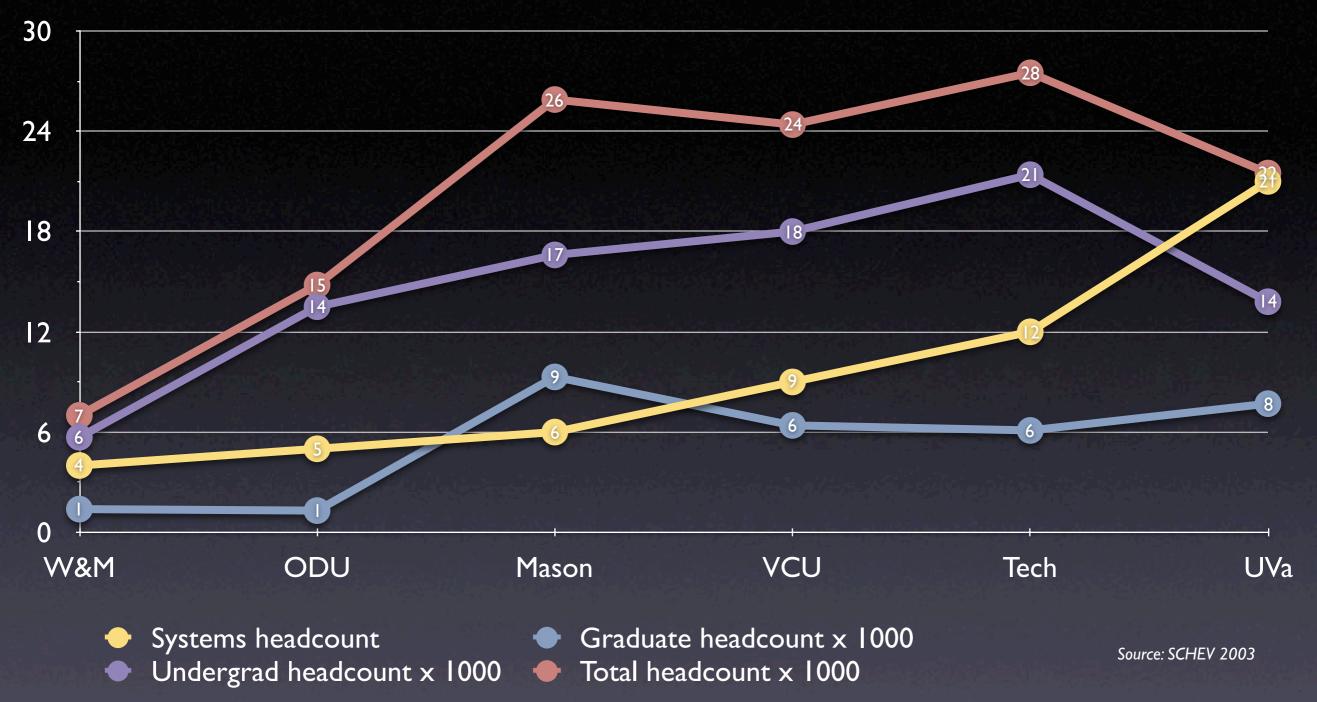


LSO Staffing

2 professional librarians 4 classified staff

- Wally Grotophorst Associate University Librarian
- Lara Bushallow Systems Librarian
- Pam Levy Application Analyst (Voyager)
- Andrew Sikorski -PC Installation/repair
- Phat Le PC Installation/repair
- Chandarari Chet Systems Assistant

Staffing levels compared



Mason has no student wages or GRA assistance in Systems. While other institutions do, they are not included in this graph.

Unlike Mason, most library IT units DO NOT host ILS system or provide sysadmin services

Headcount / Systems Staff

National Rankings for Virginia Research Universities

Based on 2000 R&D Expenditures

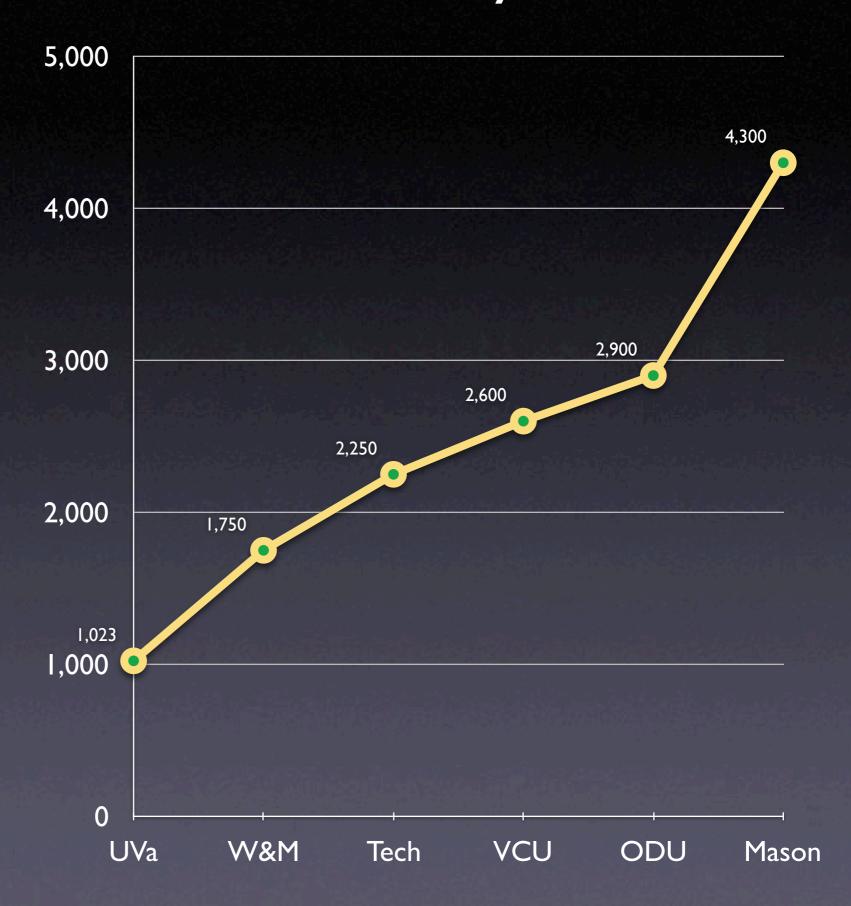
Top 100:

Virginia Tech 51^{ex}
University of Virginia 58^{ck}

Other Ranked Institutions:

Virginia Commonwealth Univ 106th College of William & Mary 157th George Mason University 173th Old Dominion University 180th

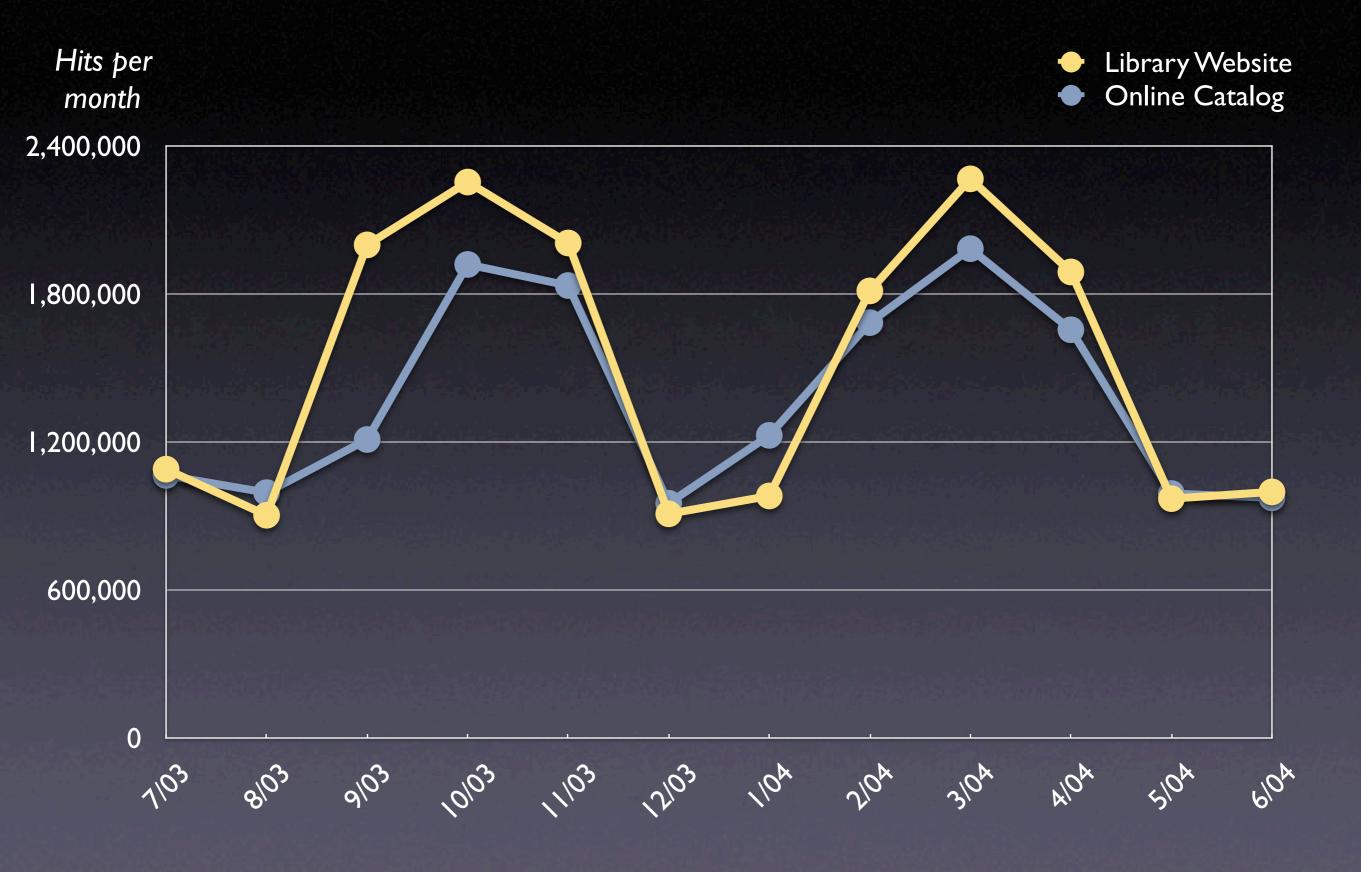
Source: National Science Foundation, Fiscal Year 2000



LSO Services

- Voyager ILS (acquisitions, cataloging, circulation, ILL)
- Locally hosted web-based services (e-reserves, e-journal finder, ILL requests, MyLibrary@Mason, static web pages)
- Authenticated off-campus access to restricted content
- Mason campus directory
- OpenURL Link Resolver (MasonLink+)
- Institutional/Digital Repository (MARS)
- 220+ In-library public workstations and instruction labs
- I 50+ Staff workstations, printers, scanners, etc.

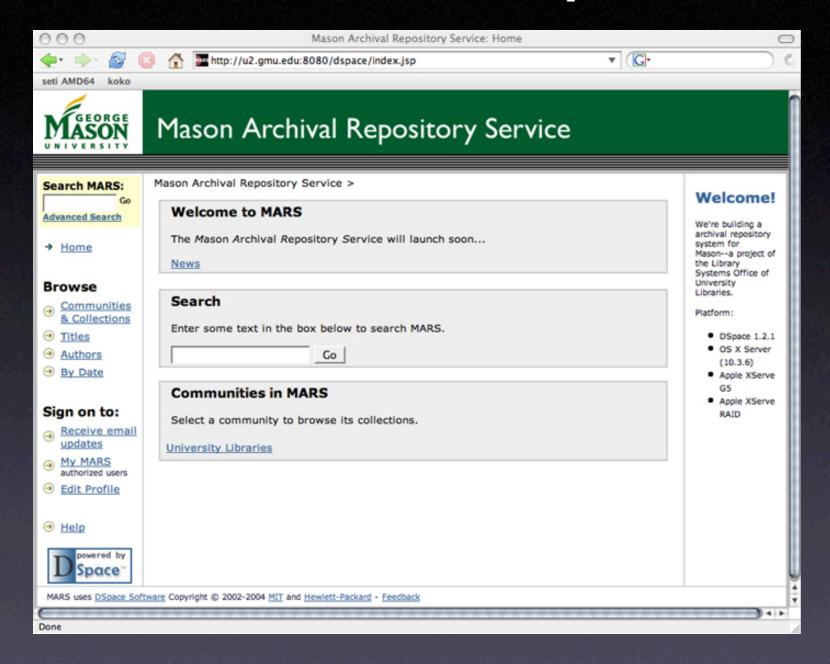
Web Visitors



LITA Top Trends - 2004

- Institutional Repositories
- Personal Search Software
- Federated Searching
- RSS
- RFID
- Copyright
- E-Resource Management

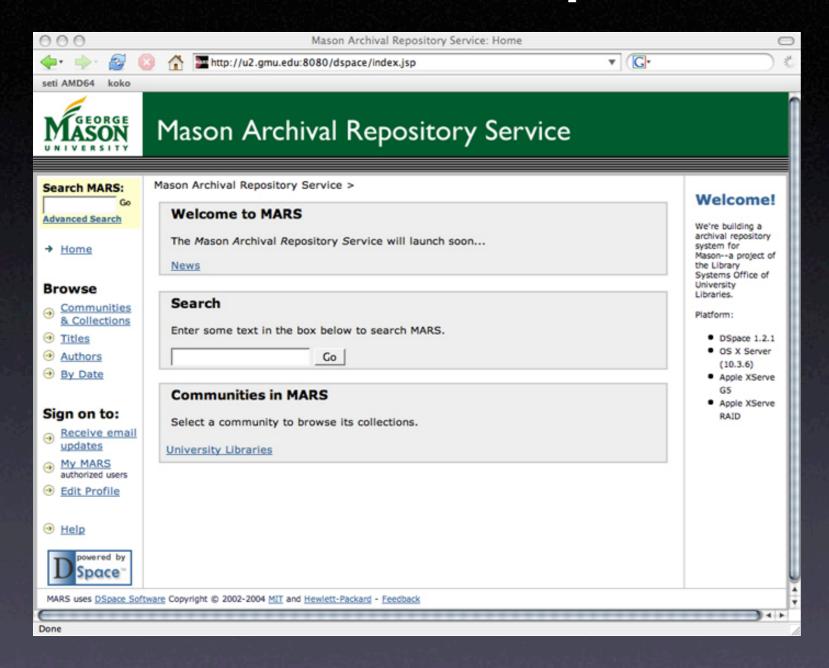
Institutional Repositories



An institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.

-- Clifford Lynch, CNI

Institutional Repositories



University Libraries is building a digital repository service for Mason. Operational today, we anticipate an early 2005 roll-out.

Mason is one of the first universities to build this system using Apple OS X Server software and the XServe RAID platform.

"DSpace **captures** your data in any format--in text, video, audio and data. It **distributes** it over the web. It **indexes** your work, so others can search and retrieve your items. It **preserves** your digital work over the long term."

-- from the DSpace "30 second elevator pitch"

http://u2.gmu.edu:8080/dspace/handle/19; 🔻 🔘





Mason Archival Repository Service





vanced Search

Home

rowse

Communities & Collections

Titles

Authors

By Date

gn on to:

Receive email updates

My MARS authorized users

Edit Profile

Help



Mason Archival Repository Service > University Libraries > MARS - Digital Repository Group >

Please use this identifier to cite or link to this ite http://hdl.handle.net/1920/7

Title: Building On Our Strengths: Final F

Digital Archiving Task Force

Authors: DAPA Task Force

Keywords: Digital Archiving

George Mason University

Issue Date: 6-Oct-2004

Abstract: Final report of the Digital Archivin

Preservation and Access Task For University Libraries. Draft published March, 2004. Surveys need for di preservation efforts, evaluates te architectures, discusses library-rel processing required. DSpace pilot recommended. You're looking at

URI: http://hdl.handle.net/1920/7

Appears in Collections: MARS - Digital Repository Group

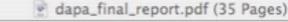
Files in This Item:

File Size Format

dapa_final_report.pdf 206Kb Adobe PDF View/Open

Show full item record

All items in MARS are protected by copyright, with all rights reserved.







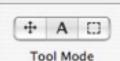












George Mason University

Building On Our Strengths: Digital Archiving, Preservation and Access

Report of the Digital Archiving, Preservation and Access Task Force

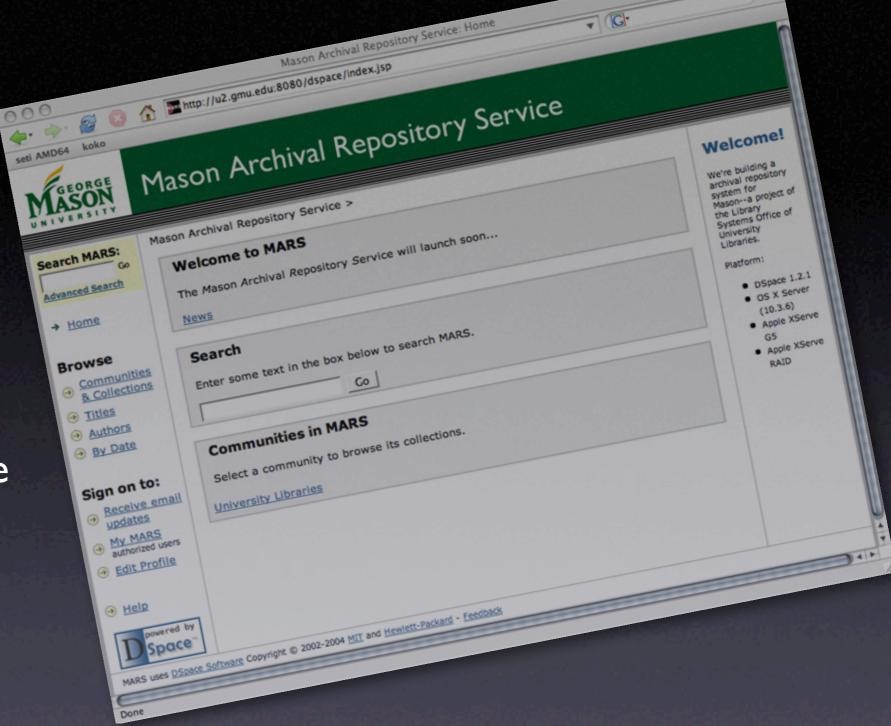
March 25, 2004

Wally Grotophorst, University Libraries, chair Dan Cohen, Center for History and New Media John Creuziger, Technical Systems Division William Fleming, University Libraries Polly Khater, University Libraries Paul Koda, University Libraries George Oberle, University Libraries Lene Palmer, University Libraries Angela Weaver, University Libraries





Thanks to compliance with OAI-MHP (Open Archives Initiative-Metadata Harvesting Protocol), items in our digital repository will be "found" by OAIster



...and other metadata havesting systems like Google Scholar and Yahoo's Content Acquisition Program.

http://u2.gmu.edu:8080/dspace-oai/request? verb=ListRecords&from=2004-08-01T14:15:00Z&metadataPrefix=oai_dc /dspace-oai/request?verb=ListRecords&from=2004

This XML file does not appear to have any style information associated with it. The document tree is shown below. OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"> // request metadatarrenx="0al_ac" verb="ListRecords"
from="2004-08-01T14:15:00Z">http://u2.gmu.edu:8080/dspace-oai/request
// ListPoorder vesponseDate>2004-12-06T14:33:30Z <id>dentifier>oai:u2.gmu.edu:1920/5</identifier> <datestamp>2004-10-05T20:36:38Z</datestamp> <ListRecords> - <record> metatrata>

<a href="m - <header> <setSpec>hdl_1920_4</setSpec> </header> <metadata> <dc:date>2004-10-05T20:36:19Z</dc:date> <dc:date>2004-10-05T20:36:19Z</dc:date> <a href="http://hdl.handle.net/1920/5/dc:identifier>http://hdl.handle.net/1920/5/dcidentifier> <dc:date>2004-10-05T20:36:19Z</dc:date> <dc:format>107171 bytes</dc:format> <dc:format>image/jpeg</dc:format> <dc:language>en_US</dc:language> <dc:subject>editorial cartoon</dc:subject> <dc:title>In Their Own Words</dc:title> destyne>Image</dc:type>

In response to an OAI-PMH request the system returns metadata about the contents of the collection...

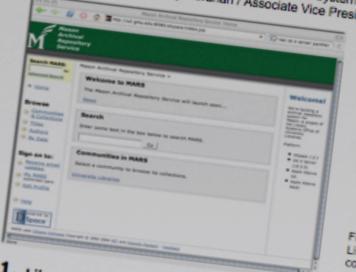
Last month, we presented a session on Mason's work in the IR area at the SPARC / SPARC Europe meeting held in Washington, D.C.



Leveraging Institutional Repository Technology to Address Archiving, Preservation and Access **Objectives Across the University**



Wally Grotophorst, Associate University Librarian / Systems John G. Zenelis, University Librarian / Associate Vice President, IT



Executive Summary:

University Libraries recognized a range of digital preservation needs across campus--many specific to the libraries but not exclusively. Bringing together stakeholders and staff expertise, a universitywide Task Force was formed, led by the Library Systems Office.

The group created a website to pull together readings and useful links, identified candidate platforms and evaluated implementation options. DSpace was selected as the platform for several reasons:

- system architecture met needs beyond open-access publishing
- open source, OAI compliant, persistent web-accessible object identifiers

Final report was shared with senior University administrators. Library was given responsibility for the initiative and new funding to cover startup hardware. A new Librarian position was authorized to assist in the management of this new library service.

1. Libraries recognition of multiple needs for digital preservation, archiving and access in the libraries and across the university.

- A previous commitment to the Center for History and New Media to explore preservation of objects in their school product archives (was characterial). •Several departments had expressed need for a publication system for working papers, individual research,
- •Multiple departments had approached library staff with need to preserve digital assets
- *Library had archiving needs of digital objects in Special Collections (film, video, oral histories, images,
- •Library needed to improve access and streamline workflows in processing digital objects A need to archive and preserve digital assets of the library (e.g., MARC database; purchased digital content,

2. Campus-wide Task Force was formed by University Libraries to explore issues and recommend solution(s)

Bringing together stakeholders and staff expertise we formed a task force drawn from library functional areas Other of the control of the control

Task force built a website (http://silo.gmu.edu/da) of readings and relevant links; formalized system Task force built a website (http://sio.gmu.edu/da) of readings and relevant links; formalized system requirements; developed a "tiers of service" model; evaluated three candidate systems (Fedora, Greenstone and Decaya and developed a pilot project test plan.

Lessons Learned Thus

- •Not all digital preservation needs can be met by DSpace, e.g., it is not very good at archiving websites.
- •Mac OS X works as a server platform and XServe RAID works well too. We believe this platform offers a way for libraries to build a low-cost, high-performance DSpace system while reducing the level of UNIX system administration expertise
- •Don't allow your IR system to be viewed as a "cost-free" digital storage service
- •DSpace can serve as a "digital warehouse" of webaddressable objects givig new options for system integration

Institutionalizing IR

The task force report was shared with the VP/CIO of the University (IT), the Provost (Academic) and the University Budget Group (Funding). The IR concept was approved by all and the library was given responsibility for what was viewed a "university project."

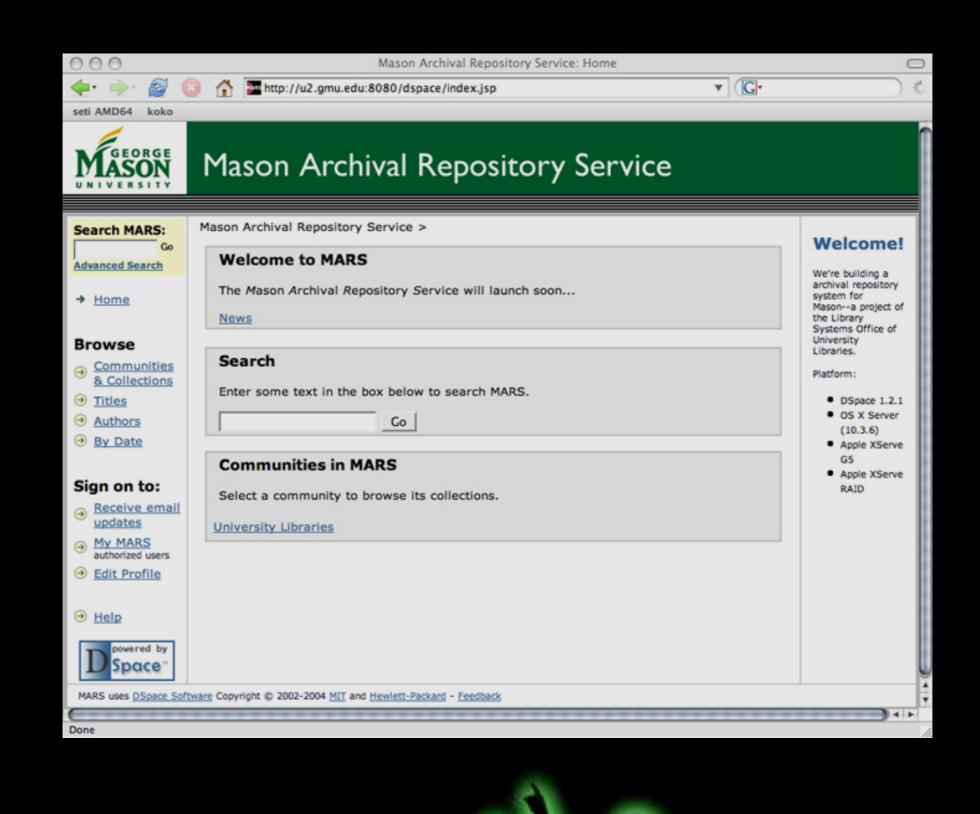
Library has selected Apple Mac OS X Server and Apple XServe hardware for our production

A new librarian position has been added to staff to manage the service and provide outreach/audience-building activity closely with our F

3. Tiers of Service

Just as we determined that DSpace could be used to satisfy a variety of digital preservation, archiving and access needs, we also realized that no one level of service would fit every audience. Based on work done at Duke University, we adopted the following tier-based approach:

- Archived Materials of significant and widespread value; complex, normalized metadata; commitment to periodic migration.
- •Preserved Materials have enduring value, but not enough to merit significant investment currently; basic metadata, supplied by conf submitters; commitment to preserve in curr



Are we ahead of the curve with this service?

Yes, but not by much...





Results 1 - 10 of about 1,550 from gmu.edu for . (0.07 seconds)

Scholar

IPDFI Simple mathematical models with very complicated dynamics

Page 1. Published in Nature, Vol. 261, p.459, June 10 1976. SIMPLE MATHEMATICAL MODELS WITH VERY, COMPLICATED DYNAMICS, Robert M. May*, Abstract. ... RM May - View as HTML - Cited by 420 Nature, 1976 - math.gmu.edu - nedwww.ipac.caltech.edu - organic.usc.edu

[PS] Web Server Workload Characterization: The Search for I MF Arlitt, CL Williamson, C Drive - View as HTML - Cited by 366 Page 1. Web Server Workload Characterization: The Search for Invarian

Version). Martin F. Arlitt Carey L. Williamson Department ... Proceedings of the ACM Sigmetrics Conference, Philadelphia, ..., 1996

cs.usask.ca - all 9 versions » [PDF] Statistical learning by 8-month-old infants

JR Saffran, RN Aslin, EL Newport - View as HTML - Cited by 325 Page 1. Institution: George Mason University | Sign In as Individual | FA Rights | Join AAAS. Statistical Learning by 8-Month-Old. Infants. ...

IPDFI Distributed representations of structure: A theory of analogical access and mapping

IE Hummel, K.I. Holyoak - View as HTML - Cited by 188 Science, 1996 - archlab.gmu.edu - dx.doi.org - ncbi.nlm.nih.gov Page 1. Psychological Review. Copyright 1997 by the American Psychological Association,

JE Hummel, KJ Holyoak - View as HTML - Cited by 168

Inc. 1997, Vol. 104, No, 3, 427-466. 0033-295X/97/\$3.00. ...

Psychological Review, 1997 - hfac.gmu.edu

Department University of California Berkeley, CA 94720. ...
1987 - classweb.gmu.edu - acm.org - cs.bgu.ac.il - vldb.org - all 1 Record The DOSTGRES Data Model? Lawrence A Rowe Mxhael R. Stonebraker. Computer LA Rowe, M Stonebraker - View as HTML - Cited by 110 [PDF] The POSTGRES Data Model

On Monday (12/6/2004) this search on Google Scholar revealed there are already 1.500+ self-archived articles on various Mason servers.



Google has developed an algorithm that makes a calculated "guess" at what it thinks is scholarly material and makes that available to the end user via the Google Scholar interface.

Google has made arrangements with publishers (e.g., database vendors) to "crawl" and index their content, then provide information at the abstract level to the end user...pulling this content out of the "invisible" or "deep" web.

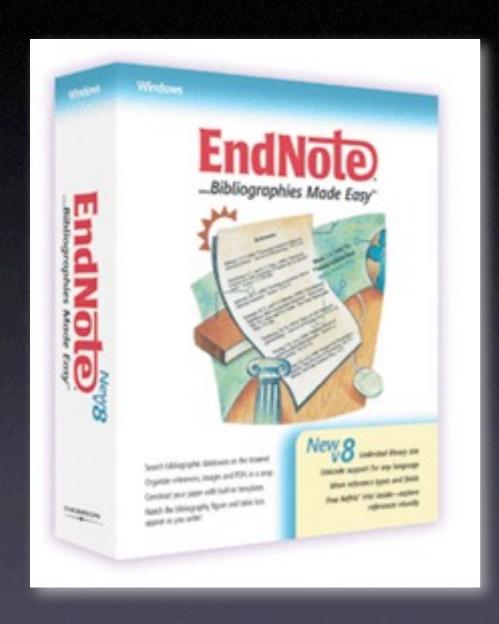
What version of content do you see? It depends on the source of the version you select and your relationship with its owner:

Full text of articles from open access journals and preprint repositories, as well as preprints on the web. For others, you should be able to get access to the full text if you or an institution you belong to has a subscription for the content. Regardless of the source, you should be able to see an abstract for any article, with the exception of those that are offline and referenced in citations only.

Interesting features:

- Citation counts
- Google requires content providers to include abstracts (at a minimum) in displayed "match" content
- Links can trigger a business transaction with the publisher
- Yahoo will likely enter this space soon (e.g., CAP program arrangements with NPR, Northwestern, OAlster, etc.)
- Exposes copyright-violations

Personal Search Software



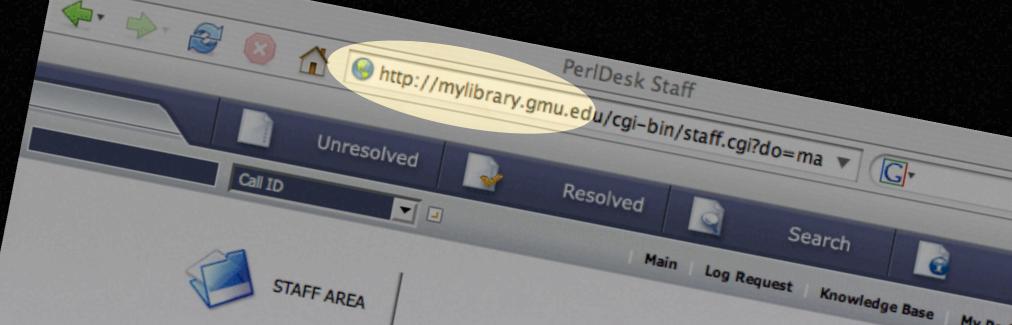
Millions of researchers, scholarly writers, students, and librarians use EndNote to search online bibliographic databases, organize their references and images, and create bibliographies and figure lists instantly.

Instead of spending hours typing bibliographies, or using index cards to organize their references, they do it the easy way—by using EndNote!

Personal Search Software

University Libraries (George Oberle and Kevin Simons) is leading an joint effort (with Technology Across the Curriculum, DoIT, and the ITU Support Center) to support use of EndNote software across campus.

- site license for all students, faculty & staff
- training sessions
- online support center



The online support center is provided by the Library Systems Office



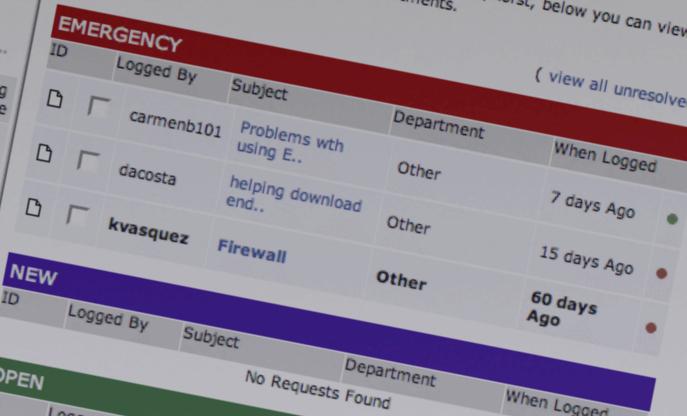
Welcome to the staff area Wally Grotophorst, here you can manage incoming requests and manage your personal settings.

Departmen	t Awaiting
302B - Matthews	Response
302H	0
Patterson 302N01 -	0
beach	0
302N - Raffel	0
130NCLC - Smith	0 10
701English - Rutledge	
Other	OP
***************************************	0 ID

PM INBOX (open) You have 0 unread private messages

Welcome Wally Grotophorst

Thank you for logging into the staff area Wally Grotophorst, below you can view overview of support requests in your departments.



OPEN ID Logged By Subject

Department No Requests Found When Logged

When Logged





Federated Searching









"The Network is the Database"

- Library Catalogs
- Citation Databases
- FullText Databases
- E-Journals
- Aggregated Content
- Open Web
- "Deep Web"
- Digital archives
- Institutional Repositories

"Noise" in the system

- Many sources of information
- Many different user interfaces
- Content / format varies across sources
- Overlapping coverage
- Undocumented gaps in coverage
- Easy to miss content or be flooded with irrelevant information
- Must deal with inevitable duplication

The Problem

- Figuring out where to begin a search is a challenge for many users, particularly undergraduates.
- Selecting the "optimal" resources requires experience with organizational schemes imposed by the discipline, by libraries and by content providers.
- There typically is no "one best path."

Solutions?

- Cross-database searching (e.g., federated searching, metasearching, or broadcast searching)
- Content linking

Federated Searching

- For the user, federated searching is a simple search box that retrieves content from a wide variety of databases and e-resource collections.
- Who does that sound like...?



But it's not like Google

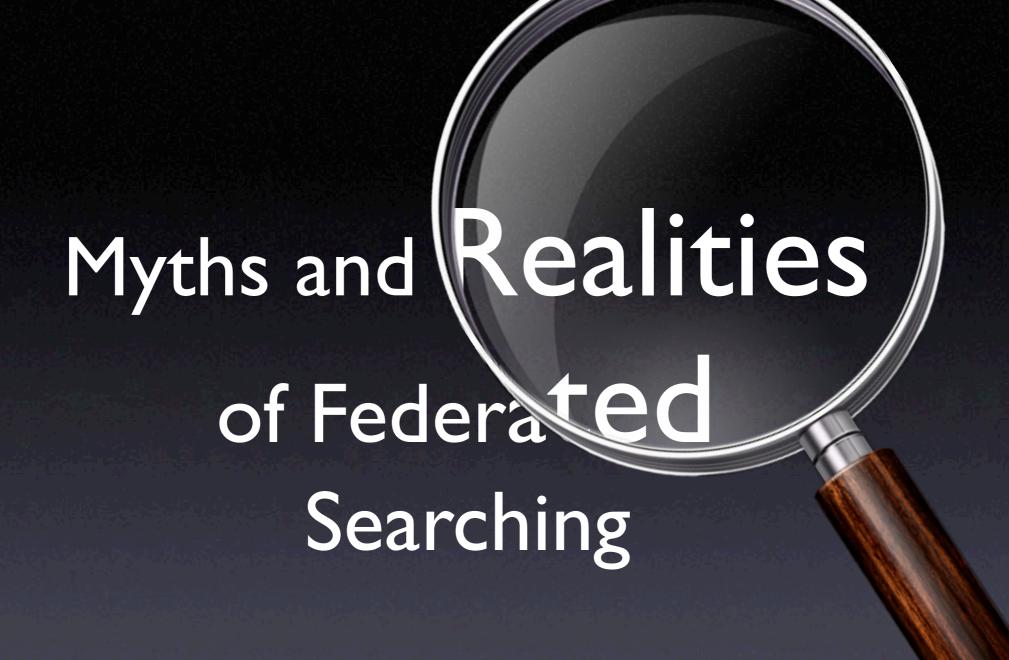
- Google sends its robots to collect data from millions of web pages in advance, so the user is actually searching a cross-file index, not the content that created the index
- Users connect to the full-text pages when they click on a hyperlink in the results set.

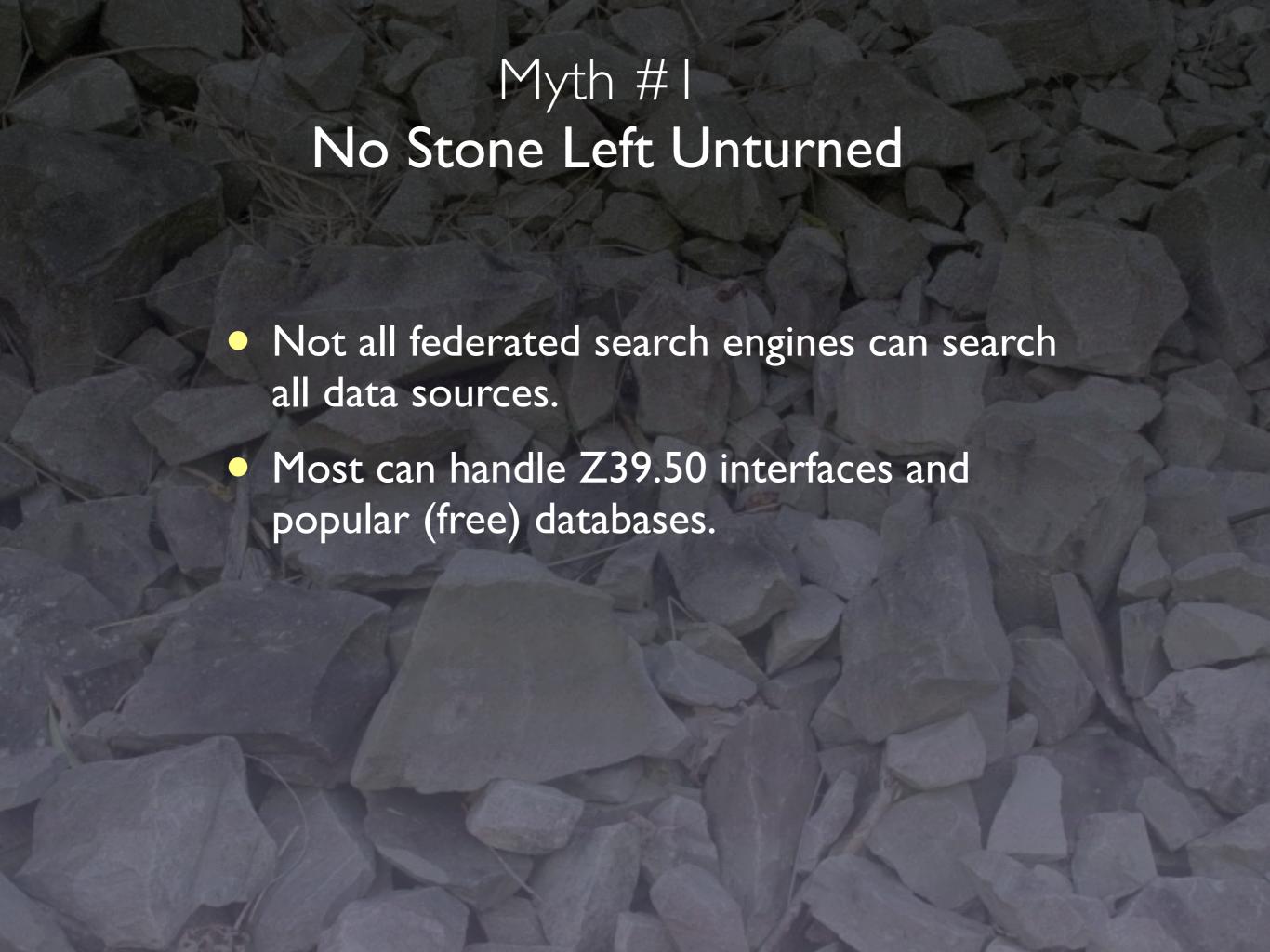
A Federated Search...

- Executes a cross-file query across citation and full-text databases that do not share a common thesaurus or index.
- A different search protocol may required for each source.

Search protocol challenges

- Some use variations of Z39.50 protocol which predates the web.
- Some use XML to identify the data elements being used
- Many leave it to the search engine vendor to figure out how to access the content (screen scraping).





Myth #2 De-duplication really works

- True de-duping is virtually impossible. The search engine would have to download all results and compare them on multiple data elements (often poorly documented)
- The limiting factor is the way in which data sources return results -- 10 to 20 at a time. Completing a true de-dupe would take hours as a typical search might return hits in the thousands...

Myth #3 Relevancy Rankings are Relevant

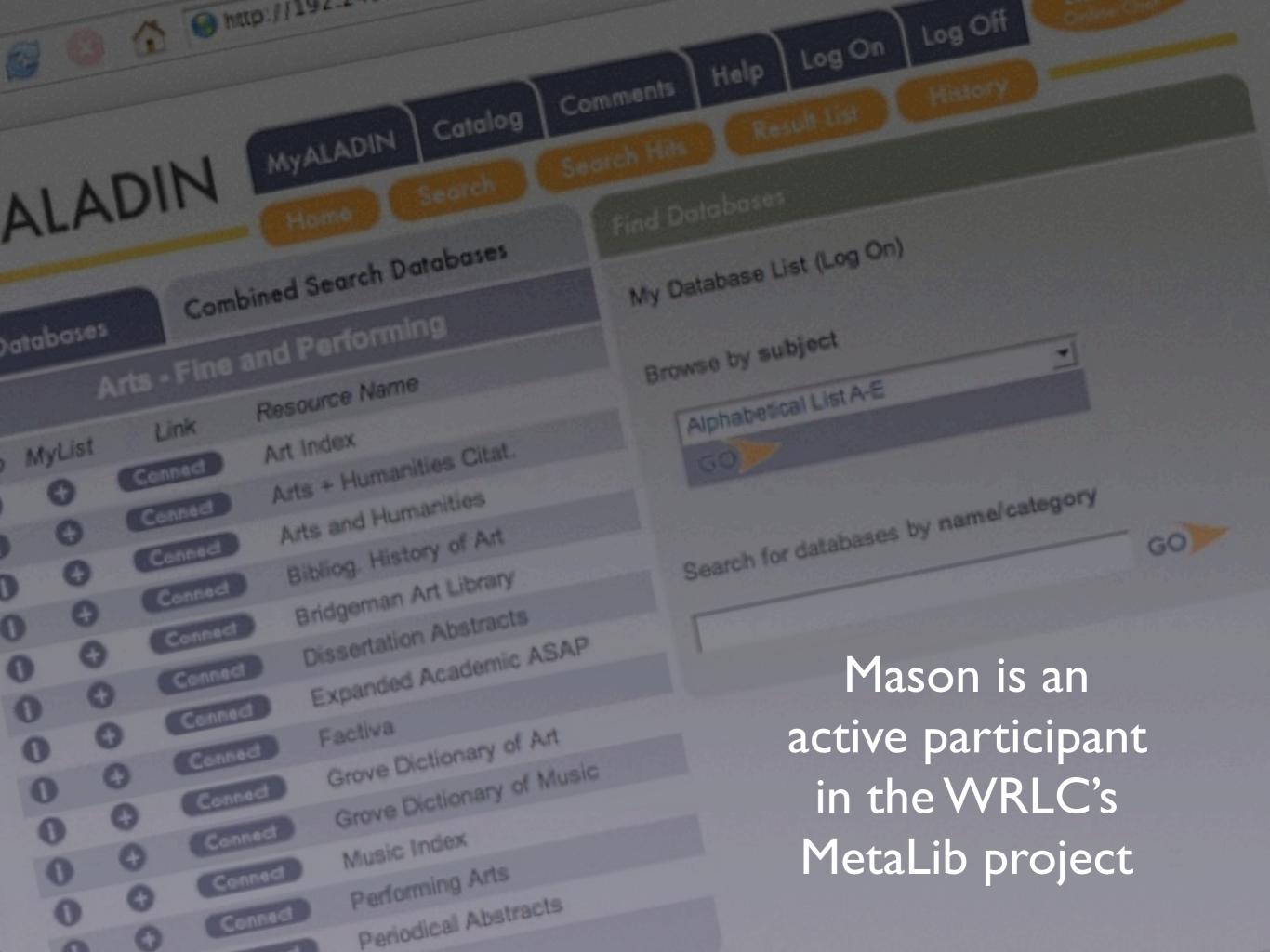
- When attempting to relevancy-rank citations, the only words you have to work with are those that appear in the citation. Often, the search word doesn't appear.
- The abstract and full-text data, as well as the indexing that content providers use to relevancy-rank their content, are unavailable to federated search engines

Myth #4 E-Resource Management is Simplified

- A federated search engine searches databases that update and change an average of 2 to 3 times a year. A system accessing 200 databases is subject to 400 to 600 updates per year--better than two a day.
- When updates are not made, access to content is lost and expensive resources go unused.
- Maintenance of this software layer is expensive.

Our take...

- University Libraries believes that federated search engines show promise but we do not feel they are ready for anything more than experimental use.
- We are *monitoring developments* in this area closely and positioned to expand our use of the technology as it becomes useful to do so.



Content Linking

Federated searching tackles information overload on the front end--attempting to improve the efficiency of the discovery process.

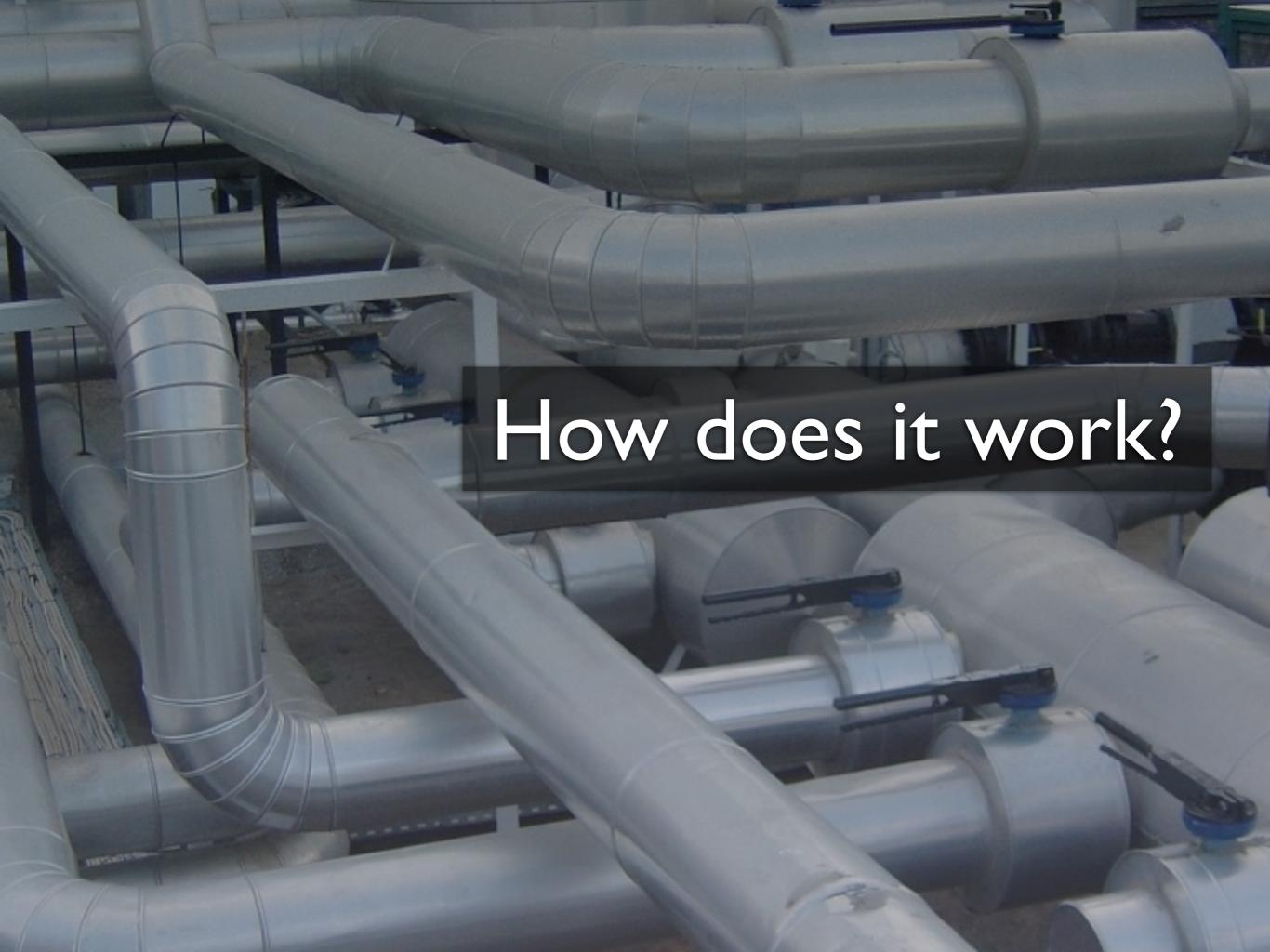
Content linking focuses on the back end--you've found a useful source but how do you know that it is the most complete version available? And is there "related" information that you might be missing?

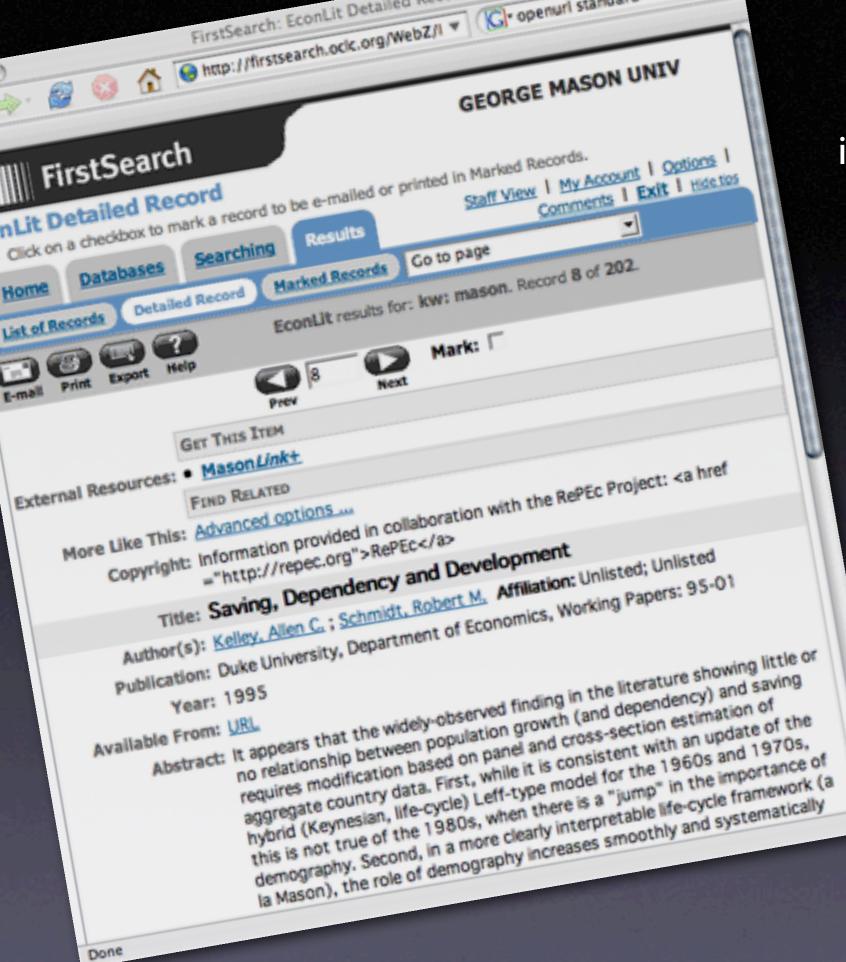
OpenURL

```
http://www.mysrv.org/menu?
id=doi:10.111/12345&
genre=article&
aulast=VVeibel&aufirst=Stu&ISSN=35345353
&year=2001&volume=14&issue=3&spage=44
&
pid=2829393&
sid=OCLC:Inspec
```

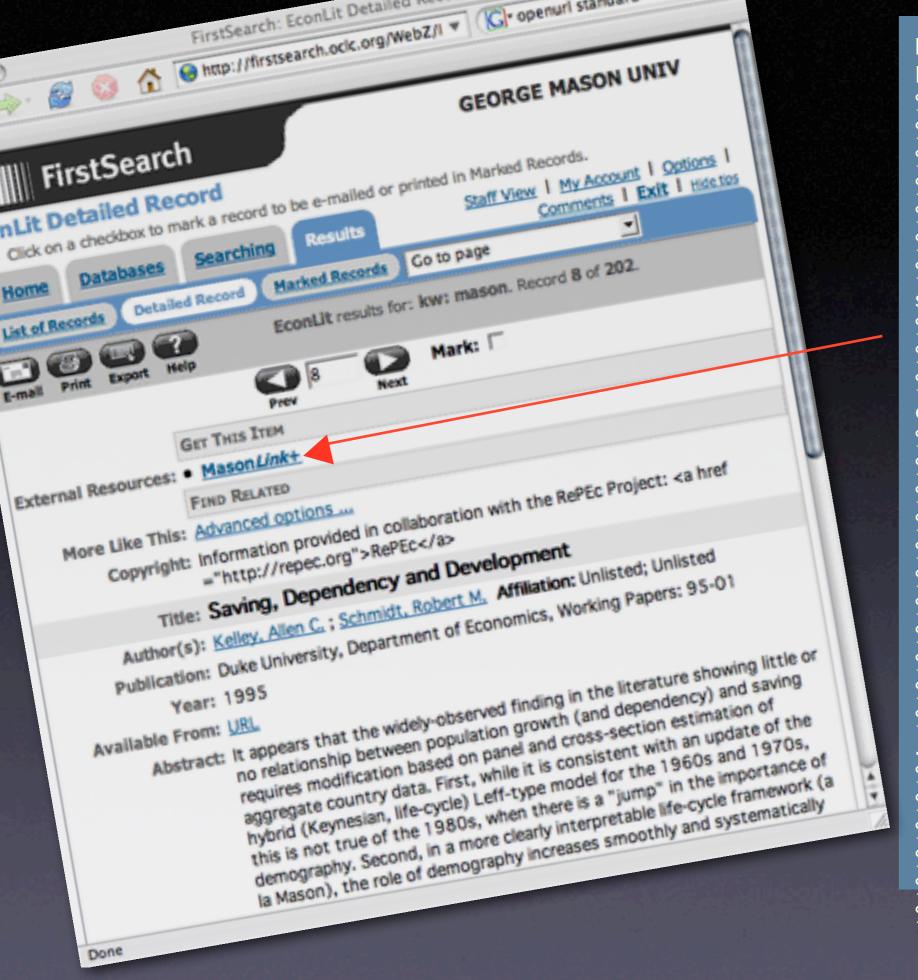
Sample OpenURL

- This service relies on the OpenURL standard (Z39.88-2004)
- OpenURL is a syntax for encoding metadata into a URL so that it might be passed between systems via the HTTP protocol.

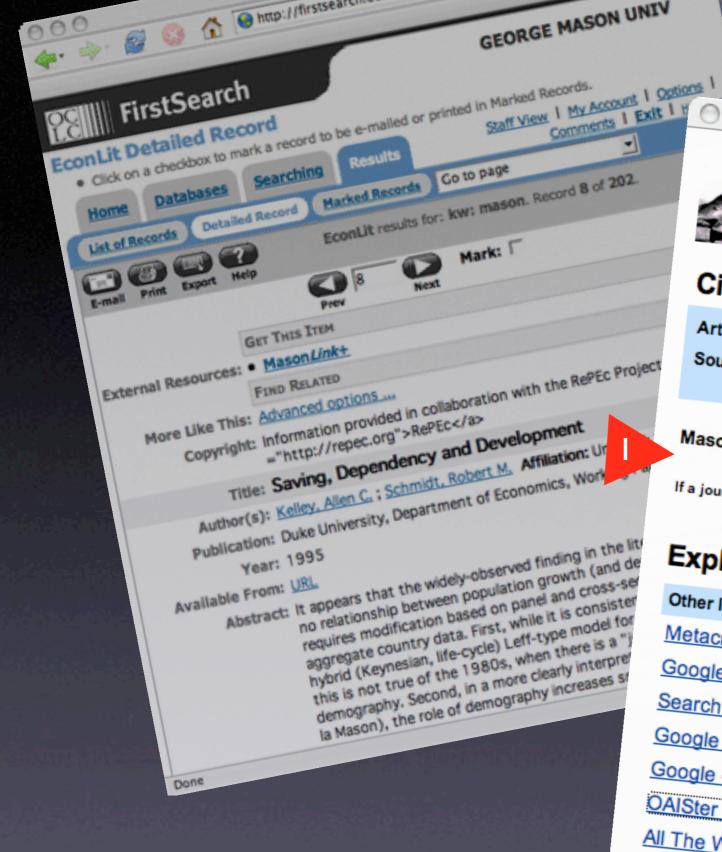




We've done a search in EconLit and found an interesting citation. We click on MasonLink+



http://129.174.55.84:8888/lfp/ LinkFinderPlus/Display??sid=FirstSearch %3AEconLit&genre=article&atitle=Saving %2C%20Dependency%20and %20Development&title=Duke %20University%2C%20Department%20of %20Economics%2C%20Working %20Papers&issue= %2095-01&date=1995&aulast=Kelley&aufir st=Allen&auinitm=C&id=doi%3A&pid= %3Caccession%20number %3E0698586%3C%2Faccession%20number %3E%3Cfssessid%3Esp06sw04-39299e2erws1z-q2wfin%3C%2Ffssessid%3E %3Cedition%3E%3C%2Fedition %3E&url ver=Z39.88-2004&rfr id=info %3Asid%2Ffirstsearch.oclc.org %3AEconLit&rft val fmt=info%3Aofi %2Ffmt%3Akev%3Amtx %3Ajournal&req id=%3Csessionid %3Esp06sw04-39299-e2erws1z-q2wfin %3C%2Fsessionid%3E&rfe dat= %3Caccessionnumber%3E0698586%3C %2Faccessionnumber %3E&rft.aulast=Kelley&rft.aufirst=Allen&rft .auinitm=C&rft.atitle=Saving%2C %20Dependency%20and %20Development&rft.jtitle=Duke %20University%2C%20Department%20of %20Economics%2C%20Working %20Papers&rft.date=1995&rft.issue= %2095-01&rft.genre=article





Citation Information

Edit Citation

Article Title: Saving, Dependency and Development Source:

Duke University, Department of Economics, Working Papers; Issue: 95-01; Date: 1995

MasonLink+ is unable to make a direct link to this item.

If a journal article, try the Search E-Journal Finder link below.

Explore Related Information

Other Internet Resources

Metacrawler - Title Search

Google - Journal Title Search

Search All The Web for Journal Info

Google - Search Author

Google - Article Title Search

OAISter - Search Article Title

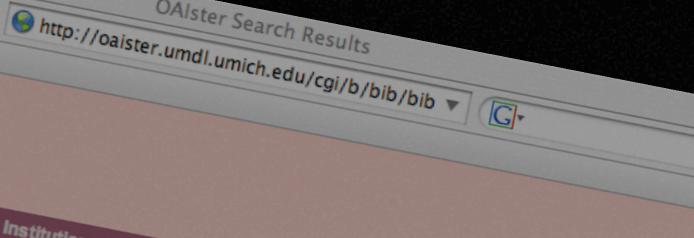
All The Web - Author Search

Publist - Search Journal Info



MasonLink+

Done



cross-section estimation of aggregate country data. First, while it is

consistent with an update of the hybrid (Keynesian, life-cycle) Leff-type model for the 1960s and 1970s, this is not true of the 1980s, when there is a "jump" in the importance of demography. Second, in a more clearly interpretable life-cycle framework (a la

Systematically over the 30 years, although the form which it takes (life-cycle versus lifetime level) also varies systematically over time. Overall, however, demographic factors accounted for a

Mason), the role of demography increases smoothly and

major portion of changes in saving across countries

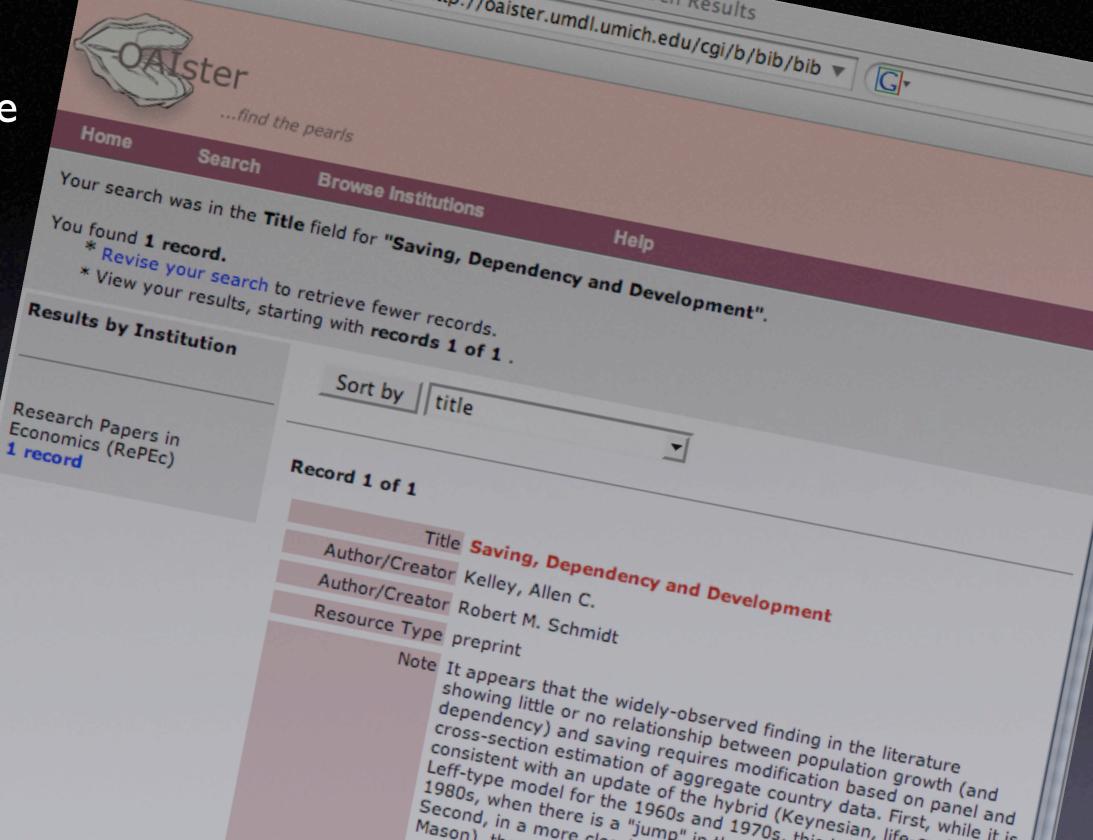
Note RePEc:duk:dukeec:95-01

URL http://www.aca

Institute

We find the "working paper" via **OAlster**

Done



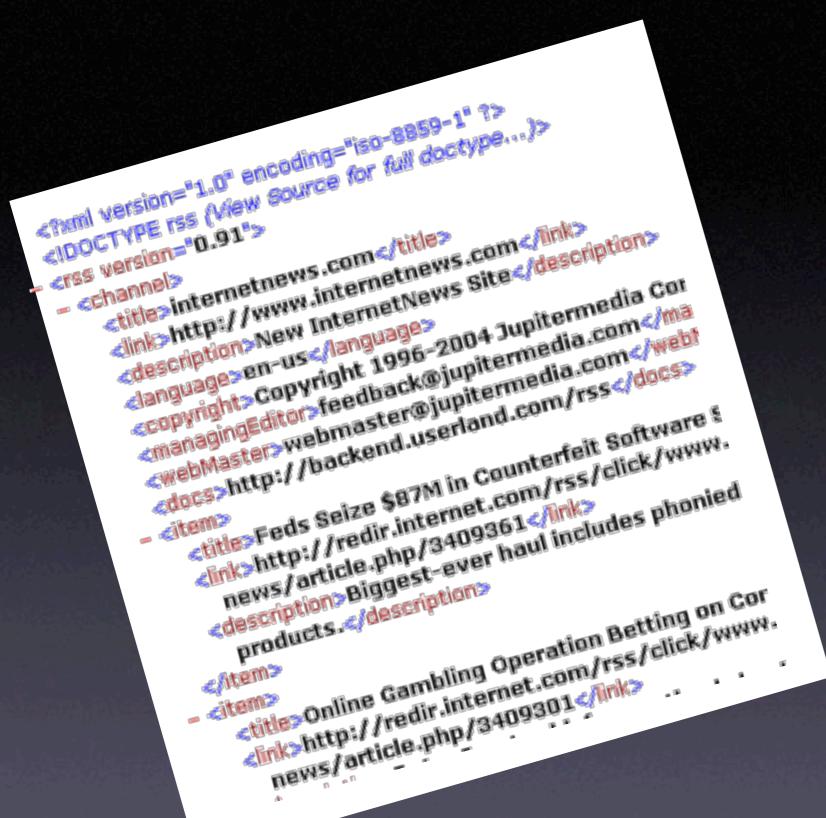
MasonLink+

MasonLink+ is an OpenURL link resolver.

It allows the user to move from a citation in one vendor's database to the full-text version of the same content in another.

It enables the university to get more value and use from the content we pay to license.

It offers the library a tool (and development platform) to deliver related but remote content

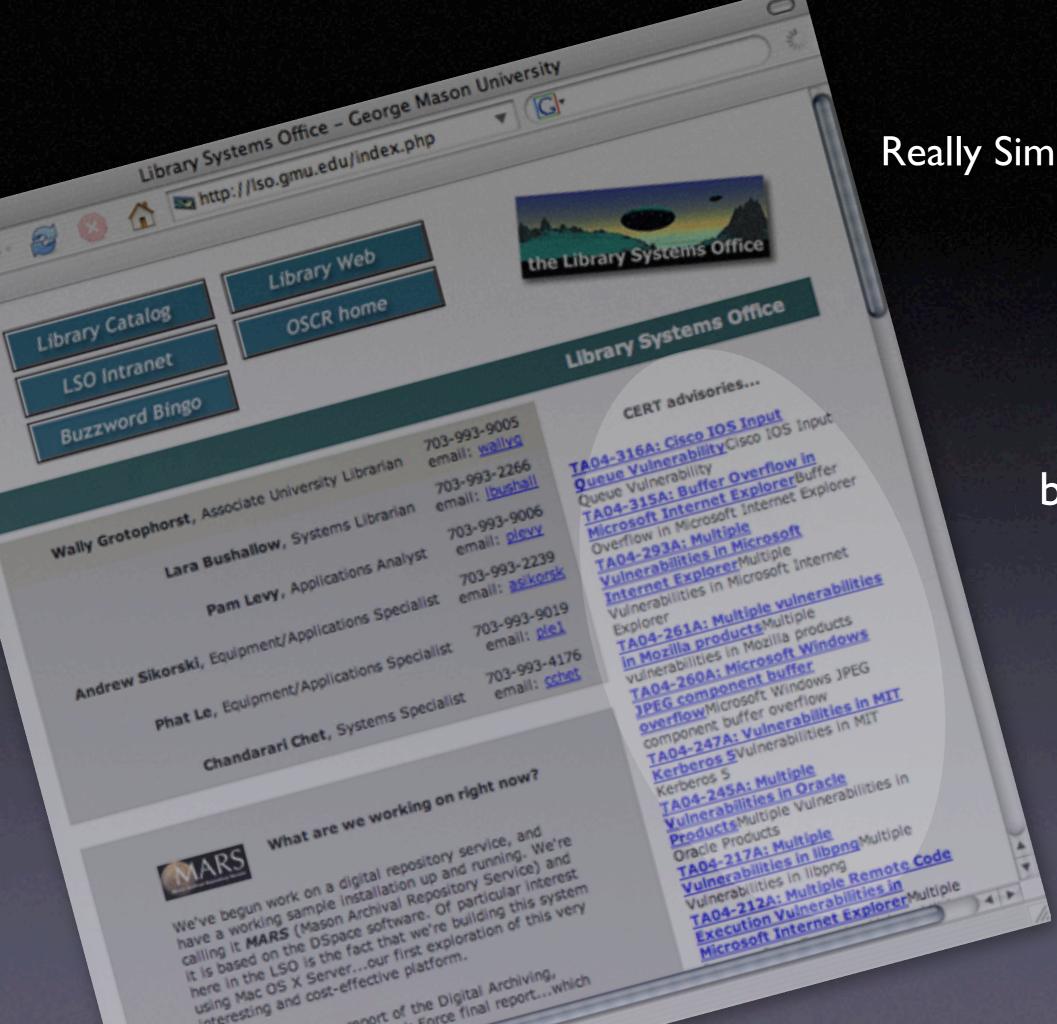


RSS

Really Simple Syndication

RSS is an XML-based format for distributing and aggregating Web content.

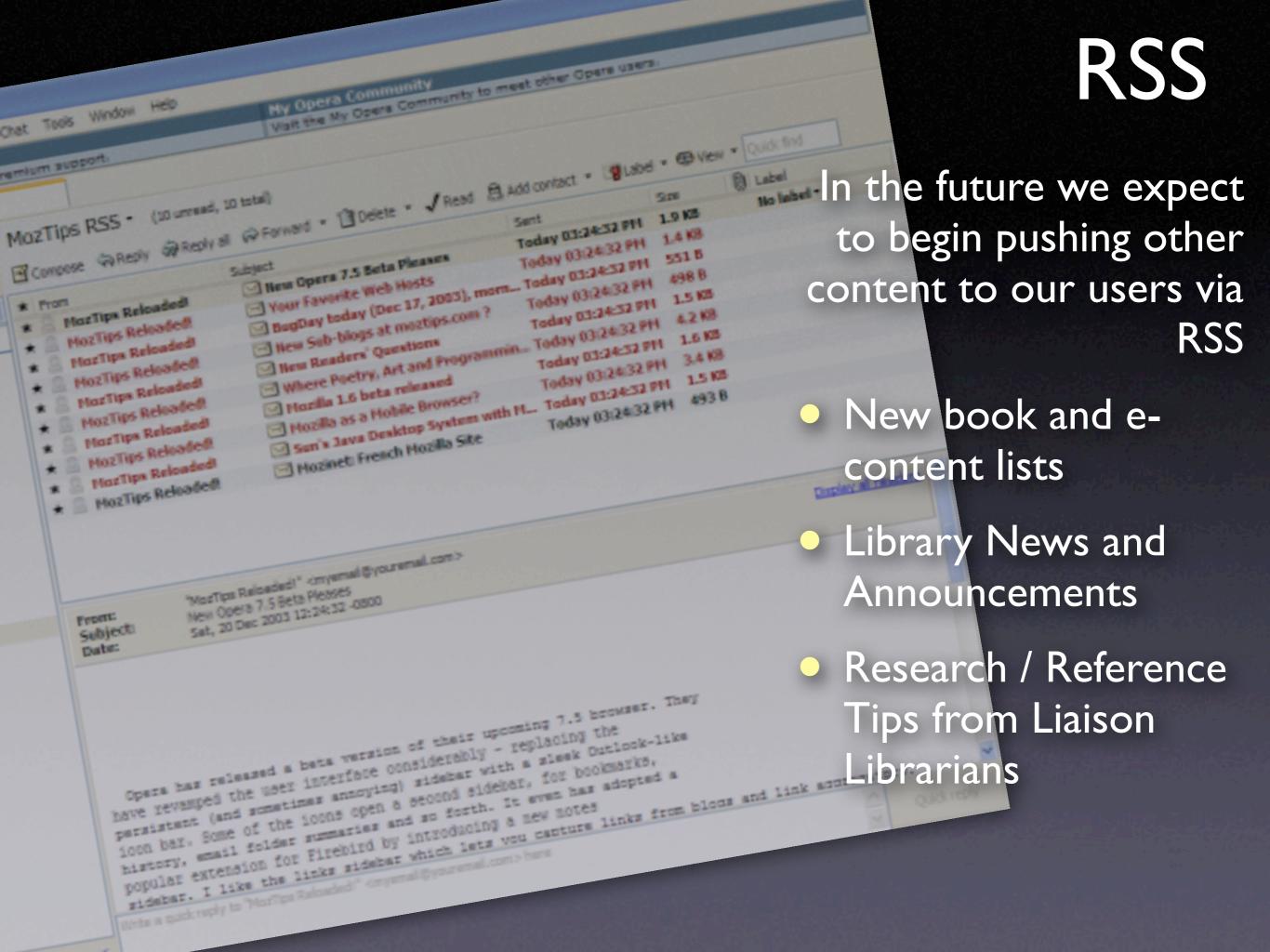
The format can be used to "feed" RSS readers or other websites (e.g., portals, blogs, etc.)



RSS

Really Simple Syndication

University
Libraries is
currently
using RSS to
build dynamic
content on
some web
pages





RFID

Radio Frequency Identification

University Libraries is not contemplating a conversion to RFID security in the near future but will consider the technology when we implement our next generation checkpoint control system.

RFID is expensive for larger libraries. e.g., RFID tags cost \$0.85 each





By 2010...

We will have to expand and improve facilities, increase systems office staff and improve the skill level and IT literacy of library staff.



We have a solid record of identifying, adapting and deploying advanced technologies... ...but to do so on a larger scale and with capacities to satisfy a more demanding audience, we will require enhanced resources--both staff and equipment.

Current facilities for mission-critical LSO systems are indequate. We recommend construction of a secure server area in the next structure built for library use.



In the interim, during 2005 we intend to co-locate LSO staff in Fenwick and Johnson Center to new facilities in Fenwick. This will:

- improve staff efficiency
- offer cross-training opportunities
- separate staff from areas where servers and storage systems are housed

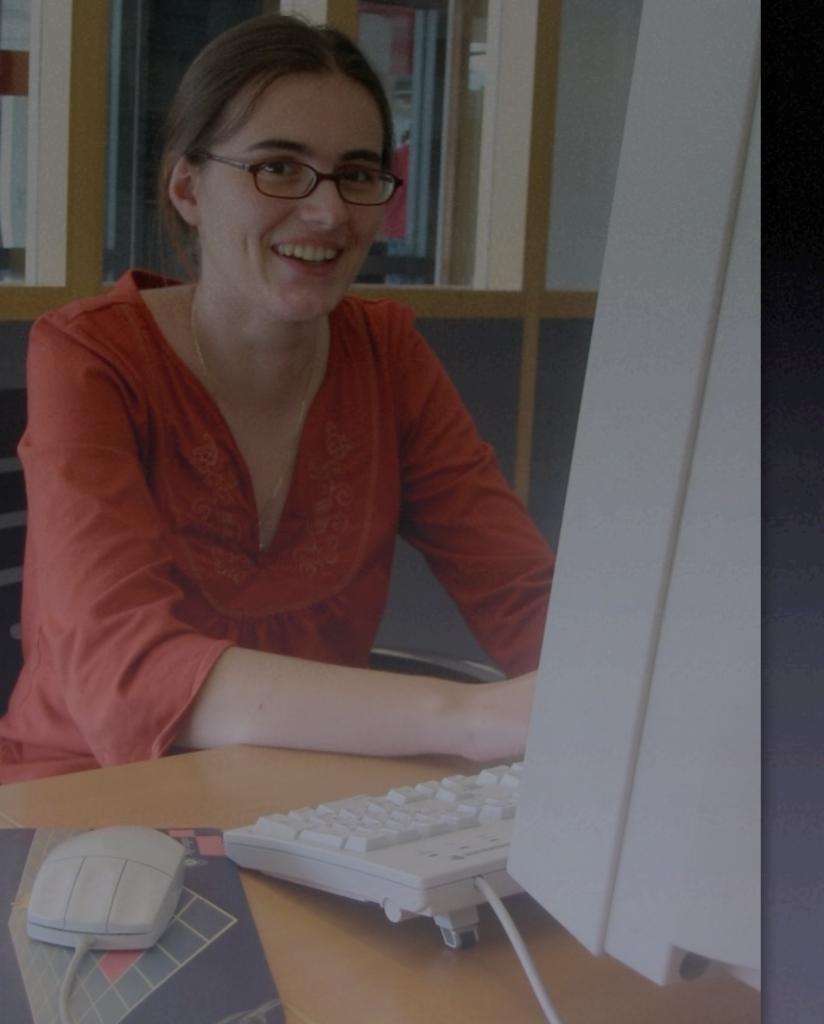
Given staffing levels within ITU and the skill sets required, we do not believe relocating our servers to an ITU-hosted area would be benefical to the university.

Hardware Funding

By 2008 all servers currently in use will have reached or exceeded their expected service lifetime.

We will need to find a funding stream to replace them.

SunFire V880 \$35,000
Three Sun E250's \$24,000
Additional storage for MARS \$20,000+



Also true for our 200+ public workstations

As we move toward 2010, we should focus more attention on providing robust, secure wireless connectivity to our inhouse users and recalibrate the number of library-provided workstations we feel we must offer.

In Summary

- Library should maintain an in-house IT operation (e.g., resist "outsourcing" either to other university operations or third-party services). The requirements of Library IT support are too specialized for generic IT assistance.
- We need to encourage central IT (ITU) to fully develop core technologies that can benefit the library (e.g., LDAP for authentication, high-speed access to off-site backup, etc.)

In Summary

• Staffing and support for Library Systems should be expected to increase between now and 2010.

Additional staff will enable University Libraries to embark on the sorts of initiatives one associates with libraries of research universities.

With workloads reduced, LSO staff can assume a greater role in staff development efforts--which can improve overall technology skills across professional staff.

• Improving LSO facilities must be an important component of new construction initiatives.

Questions?



University Libraries Library Systems Office