Document Imaging and Records Management at Berkeley

August 7, 2012
Patrick McGrath – IST Research & Content Technologies
IST Content Management Services

• Content Management Service
  • Domains
    • Document Capture Imaging (fax, scan, email)
    • Document Management
    • Business Process Management / Workflow
    • Content Integration
    • Records Management
    • Web Content Management

• Services & Technologies
  • Imagine (Lexmark/Perceptive Software ImageNow)
  • Calshare (MS Sharepoint)
  • Pantheon (Drupal)
  • Research Hub (Alfresco)
  • Box (those guys)
  • Google (those other guys)
Annual paper consumption Worldwide...

20 Million Tons

Paper is, and will continue to be a critical component of business transactions.

... from a 2010 study
Paper Usage at UC Berkeley

Q: How many pages a year does UC Berkeley generate?

A: From OE analysis, over **60,000,000 pages** from the largest supplier.

If we consider baseless estimates of 10% additional from other suppliers paper reimbursements and outsourced printing (e.g. Kinkos)

Perhaps **67,000,000 pages purchased**?

81% recycled – above average by industry
What does 67,000,000 look like?

- **134,200** reams of paper (if all was letter)
- A stack of paper **4.77 miles high**
- A **football field** of paper **4” high**
- **336 tons** of paper
- **2110 tons of CO$_2$** in the production of the paper
What’s the total cost of 67M pages

The Cost of Paper Use

It’s no secret that paper is one of the most resource intensive goods to produce, and that law firms use A LOT of paper. But paper use also carries a heavy financial burden. Most of us are familiar with the cost of purchasing paper, but its use also requires machines to print and copy it, people to find, store, and send it, and money to ensure that it is properly disposed of once you are done with it.

Answer ALL six simple questions on this form, and find out how much paper costs your office.

1.) How many employees does your firm have? 12000

   • I don’t know, use Seattle Industry average of 36 employees per firm

2.) How much paper does your office use (enter data in ONE box only)?

   Per Month
   Or

   Per Year
   Or

   # of Sheets
   # of Reams
   # of Boxes

   67,000,000

   • I don’t know, use industry average of 30,000 sheets per employee per year

3.) What does your paper cost to purchase (enter data in ONE box only)?

   Per ream
   Or,

   Per box
   Or,

   • I don’t know, use average cost of $3.50 per ream

   $3.29

4.) Approximately what percentage of your firm’s disposed paper is recycled? 81%

5.) Approximately what percentage of your firm’s disposed paper is shredded? 1%

6.) Approximately what percentage of your firm’s disposed paper is thrown away? 96%

   • I don’t know ALL of these percentages, use average Seattle firm percentages
## What’s the total cost of 67M pages?

<table>
<thead>
<tr>
<th>Itemized Costs</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Your firm’s annual purchasing costs</td>
<td>$441,000</td>
</tr>
<tr>
<td>2. Your firm’s annual paper use cost (including labor, and equipment purchase and maintenance)</td>
<td>$5,344,900</td>
</tr>
<tr>
<td>3. Your firm’s annual paper disposal costs</td>
<td>$1,713,600</td>
</tr>
</tbody>
</table>

**Question Notes**

1. Average calculated from Resource Venture data
2. Paper calculations based on 500 sheets per ream, and 10 reams per box
3. Average cost per ream based on retailer data.
4. Average recycling percentage is 79%
5. Average shredded percentage is 20%
6. Average percentage thrown away is 1%

**Costs**

1. Annual purchasing costs based on cost per sheet times the number of sheets used.
2. Annual use cost is calculated as 12.12 times the purchase cost. This is based on “Waste Prevention in Business: An Overview of Waste Prevention Potential, Benefits, and Barriers” by the Oregon Department of Environmental Quality. This is considerably less than the estimate of a Minnesota study that paper use costs are up to 31 times greater than the initial purchase price.
3. Annual disposal quantities calculated based on average paper disposal per employee rates from California Statewide Waste Characterization. Costs are calculated by determining the cost per pound to dispose, recycle, and shred ALL of the paper disposed of annually by an average Seattle firm of 36 employees. Current City of Seattle disposal rates are included. Taxes are included in recycling and garbage calculations. Shredded paper is assumed to cost $0.20 per pound.

---

About **$7.5M** a year for “Front End” printing and disposal
Cost of Managing Paper

Each document costs you...

$20 in labor to file

$120 in labor searching

$1 per document to store

$5 per document to retrieve

...and on average, 5% of documents are lost

25 hours to recreate

... from 2010 study
If EMC’s numbers hold true, what’s the impact at UCB?

<table>
<thead>
<tr>
<th>Total Docs</th>
<th>Pages purchased by UCB Annually from OfficeMax (ref OE 2010)</th>
<th>81,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated pages ordered from other sources</td>
<td>7,320,000</td>
</tr>
<tr>
<td></td>
<td>Estimated pages received from external sources</td>
<td>5,000,000</td>
</tr>
<tr>
<td></td>
<td>Estimated Total Pages Managed Annually</td>
<td>73,320,000</td>
</tr>
<tr>
<td></td>
<td>Estimated Total Documents Managed Annually</td>
<td>18,330,000</td>
</tr>
</tbody>
</table>

A) Stored Locally

- 6% of 918,500 documents cannot be disposed of immediately
- Filing Costs / Document: $20
- Transportation Costs / Mail / Document: $1
- 5 Year Storage Costs: $2
- Total: $21,079,500 Cost to file & store documents

B) Retrieved

- 5% of 45,825 documents need to be referred to later
- Search: $120
- Retrieval: $5
- 5% Documents that are lost and recreated: $1,250
- Total: $8,592,188

C) Destroyed/Archived

- 45,825 Documents
- Long term archiving costs for 5% of documents: $20
- Retention Review and Destruction Costs for 96% of documents: $25
- Total: $22,683,375

Total: $52,355,063

About $52M per year. Clearly we need to look at the numbers and the opportunities.
Imagine Service
Document Lifecycle

ImageNow helps manage every step of the document lifecycle

Capture
Collect any file format from any location.

Process
Route content automatically.

Collaborate
Work together. Smarter.

Access
Get the precise information you need.

Protect
Secure, retain and comply.

... and Analytics
Capture - Infrastructure

Capture Options:
- scanning
- faxing
- email
- print to “tiff”
- Import electronic docs
- web services

Indexing Options:
- Manual entry
- Screen Scraping
- OCR/Content Search
- Barcodes
- Database lookup
- Scripted Indexing

ImageNow Object Repository
Capture – Implementation

Original Document

<table>
<thead>
<tr>
<th>Description</th>
<th>Catalog No.</th>
<th>Quantity</th>
<th>Unit Price</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low fat milk</td>
<td>101</td>
<td>200 gls</td>
<td>20</td>
<td>$4,000</td>
</tr>
<tr>
<td>Low fat yogurt</td>
<td>102</td>
<td>400 gls</td>
<td>10</td>
<td>$4,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$9,000</td>
</tr>
</tbody>
</table>

Imagine Queue

Document in Imagine with meta data

- Invoice Number: 06312486
- Invoice Date: 09/09/2009
- Customer: UC Berkeley
Process – Screen Scraping

PeopleSoft - BFS  Metadata  Imagine
Document Management Tasks

Annotations

Doc. History

Process – cont’d

Thumbnails
Why an Enterprise Approach?
The problem with managing information assets today

1. Siloed tools and systems for creating, managing, and sharing documents
2. Massive duplication within and across repositories
3. Reactive, one-off approaches to e-discovery
4. Disconnect between retention schedule and how the business works
5. Everyone keeps everything forever
The problem with managing information assets today

Organizations are finding that records management (RM), as typically practiced, is untenable

- Retention schedules are too complex for business users to follow or for technology to support.
- Records management practices worked for paper in boxes managed by RM specialists; it doesn’t work for electronically stored information (ESI) managed by end users.
- Most ESI is retained forever, on multiple systems, and in multiple media.
The problem with managing information assets today

E-discovery is getting more difficult and expensive

- Most organizations adopt a reactive, one-off approach.
- Millions are spent grinding through piles of documents for a single legal matter... then they turn around and do it all over again (with the same documents) for the next legal matter.
- E-discovery tools have become much more sophisticated for searching, but fail to address the process management and workflow needs of managing collections across large numbers of custodians.
The problem with managing information assets today

Storage management is becoming a chronic problem for IT

- The cost of storage for most organizations will become one of the single largest expenses within the IT budget in the next five years (2nd to labor)...even though cost of raw storage is decreasing

- Meeting performance and business continuity SLAs is expensive and difficult (if not impossible)

- Tactics such as file compression and migration to tiered storage only provide a short-term solution
Why is eRM a Priority Now?

- **Savings and efficiencies**
  - Very costly current processes
  - Redundancy of space and effort, distributed workforce
  - Mission critical, time sensitive
  - Improves efficiency and productivity
  - Supports management decision making

- **Institutional memory (knowledge of the records)**
  - Aging workforce
  - Knowledge of filing schemes, data policies

- **Massive organizational change**
  - Current large investment in campus systems
  - Organization, reorganization, new reliance on partners
  - New processes, forms

- **Risk Management and Compliance**
  - Controls the creation and growth of records
  - Ensures regulatory compliance, Safeguards high value university assets (e.g. sale of Sargent Johnson wooden panels), Safeguards vital information in case of disaster,
  - Automated and consistent records retention and disposal
  - eDiscovery, minimizes litigation risks

- **Parallels between academic and administrative Records Management and Archiving functions**
In 57% of organizations, recent events such as the BP disaster, PG&E, the Toyota recalls, and the banking crisis have made senior management “more” or “much more” conscious of risks to the business.

83% of respondents feel that the importance of records management has increased in the last 2-3 years, considerably so for 40% of them.

Over the last 2 years, budgets for records and information management have increased in 43% of organizations and decreased in only 13%. They are expected to increase further in the next 2 years.

If challenged, 37% of respondents are not confident that their electronic records have not been modified, deleted or inappropriately accessed.

For 31% of respondents (excluding 45% “Don’t Knows”), a lack of complete electronic information has been an issue with regulation authorities, and for 34% it has been an issue in a court case.

In 26% of organizations, undeleted records beyond their retention period have affected a court case, twice as often weakening it rather than strengthening it.

20% of respondents have more than 50 discovery cases a year for regulatory or legal issues. 13% deal with more than 120 cases per year.

24% have had their records keeping policies questioned in court and 25% have sent a senior IT person to testify to the validity of their electronic records.

26% have undiscriminating policies on deletion of all emails, 23% keep everything just in case and 31% have no policies or non-enforced policies. The remainder either manually or automatically declare important emails as records and delete the others.

For organizations with no RM system, a legal discovery process will take an average of 25 days. For those with a system the average is halved to 12 days. For comparison, a legal discovery process across paper records takes an average of 19 days, with 28% of organizations taking more than a month.

44% of those with an RM system consider electronic-discovery to be 3 or more times as efficient as paper discovery.

60% of e-Discovery system users consider it to be at least twice as efficient as a basic electronic search, with 49% achieving payback in 18 months or less, and 72% within 2 years.

Planned spending in the next 12 months is up in all areas of records management, particularly Enterprise Search and physical records management.
Records Management Ecosystem

RM Ecosystem Overview

Retention & Destruction Policy Definition

Records Management Practices
- Clear Understood Adhered to Embedded Monitored

- Document Creation
- Amendment & Versioning

- Processing
- Control Destruction

Storage and Retrieval
- Preservation
- Permanent Archive

Active content and transaction flows

Broader Records Management Activities
Proof-of-Concept Project/s

- Objective: Assess some representative needs, experiment with an eRM package to make recommendations for wider campus eRM approach.
  - Determine if and how an enterprise-wide records management system (eRM) can improve UCB’s record-keeping practices by creating functional-based best practices and standards.
  - Determine if and how an eRM can improve efficiencies and reduce costs and risks associated with record creation and recordkeeping and discovery.
  - Determine if and how an eRM can streamline discovery tools and processes.
  - Gain UCB support for an eRM initiative.

- Project Reps
  - Human Resources - Rick Mena, Nelcy Dwight
  - Chancellor’s Office - Cindy Major, Liane Ko, Julie Oshiro
  - Archivist Office - David Farrell, David De Lorenzo, Josh Schneider
  - IST - Patrick McGrath, Ronnie Ong, Ian Crew

- Deliver reports back to IDC, campus:
  - Costs, level of effort
  - Skills required
  - Potential savings and other benefits
  - Assurances required for viability
  - Recommendations for next steps
PoC Findings to Date

- Massive culture change issues
  - Distrust of electronic formats, electronic signatures, etc. “MUST... HAVE... PAPER....”
  - Doing the best with what we have. Departments short staffed for effective compliance.
  - Limited knowledge of RM practices. Little big-picture thinking. Differences between policy and local practices.
  - Don’t know or don’t trust the System or Department of Record

- Physical location of documents across the campus are very distributed, inconsistently managed, and stored in areas that are present great risk
  - Physical security, fire, redundancy, environmentals
  - Lost paperwork
  - Inconsistent use of archives, significant loss of metadata/context when using already electronic documents
  - Significant levels of document duplication/replication

- Effort differences
  - Records Ingest due to document scanning can take longer than putting paperwork in folders and filing in the basement. However, in many cases this can be addressed.
  - Managing the files, disposition, long term archiving presents HUGE benefits.
  - Access to information presents HUGE benefits as does consistent treatment of documents to policy.

- Records often hard to get a hold of.

- Significant potential overall savings in floor space, environmentals, staff, interoffice mail, time to information, compliance to controls/policy/security
  - Significant opportunity at campus/system level.
Research Hub based on Alfresco

Technical Architecture

Campus Applications

- Research Hub (Alfresco Share)
- cSpace
- Drupal
- bSpace

Content Hub (Alfresco ECM Repository)
- ECM Workflow, Transformation Service, CMIS, Alfresco API

Core Services (LDAP, SHIB, CAS, Other)

Other Services
- DuraCloud
- CDL/UC3
- eScholarship
- Bamboo Services
- Google
- Amazon
- GenBank
Opportunities

- Parallels between academic and administrative
- Cost Savings
- Skills and Partnerships
- Pilot Projects
  - CCRC
  - Human Resources
- Awareness, Education, Services

Partnerships

- IDC
- IST-RCT
- UARC
- CDL-UC3
- Office of Ethics, Risk Mgmt and Compliance
- Office of Sustainability
- HR, CCRC

Issues

- Feasibility
- Paper and eRM hybrid models
- Cost / Business Models including UC3
  - Long term funding
- Transactional content management / app relationships
- “Fixed Term Archiving”
  - Deaccessioning
UCB Records Management PoC
Screen Shots

- Using Alfresco / Records Management Module
Hybrid A - Test Case 2 (Incoming paper letter).pdf

Identification and Status
Name: Hybrid A - Test Case 2 (Incoming paper letter).pdf
Unique Record Identifier: 2012-000001230
Record Category Identifier: 2012-000001111
Title: Incoming Letter
Description: The "Re:" Field
Owner: (None)
Declared: Yes
Date Declared: Fri 01 Jun 2012
Declared By: jschnaid
Author: Lawrence Pitts

Creator: jschnaid
Created Date: Fri 25 May 2012 11:02:42
Modifier: System
Modified Date: Fri 01 Jun 2012 10:30:00
Size: 37 KB
Mimetype: Adobe PDF Document

Record
Record Type: PDF Record
Originator: Lawrence Pitts
Originating Organization: UCOP VP Academic Affairs
Date Filed: Fri 01 Jun 2012
Publication Date: Thu 03 May 2012
Location: (None)
Media Type: (None)
Disposition Config Report

**CCRC**

Path: /
Record Series Identifier: 2012-0000001110
Description:

**Archival**

Path: /CCRC/
Record Category Identifier: 2012-0000001111
Disposition Authority: UC Records Retention and Disposition Schedule, CCRC Internal

Disposition Schedule:
1. Cutoff Immediately.
2. Transfer custody of digital records to University Archives after 10 years.
3. Transfer to Merritt after 5 years.
4. Retain for 5 years.

**Incoming Correspondence**

Path: /CCRC/Archival/
Record Folder Identifier: 2012-0000001288

**Outgoing Correspondence**

Path: /CCRC/Archival/
Record Folder Identifier: 2012-000000134

**Temporary**

Path: /CCRC/
Record Category Identifier: 2012-0000001113
Disposition Authority: UC Records Retention and Disposition Schedule, CCRC Internal

Disposition Schedule:
1. Retain for 5 years.
2. Destroy 5 years from cut-off date.
3. Cut off immediately.
Archival
Record Category Identifier: 2012-0000001111
Disposition Authority: UC Records Retention and Disposition Schedule, CCRC Internal
Disposition Instructions: Retain for 5 years, then transfer to Merritt for 5 years, then transfer to University
Vital Record Indicator: No
Created by: Josh Schneider  Modified on: Fri 25 May 2012 11:02:16

Temporary
Record Category Identifier: 2012-0000001113
Disposition Authority: UC Records Retention and Disposition Schedule, CCRC Internal
Disposition Instructions: Retain for 5 years, then destroy.
Vital Record Indicator: No
Created by: Josh Schneider  Modified on: Fri 1 Jun 2012 15:09:03

To Be Filed
Record Category Identifier: 2012-0000001117
Disposition Authority:
Disposition Instructions:
Vital Record Indicator: No
Folder Disposition Config

Records Management Site
Location: File Plan > CCRC > Archival

Records Category

Metadata
- MAXUID: 0
- Name: Archival
- UIDVALIDITY: (None)
- Title: (None)
- Description: Permanent records.
- CHANGETOKEN: 7480b6c-3e6c-43ee-bc04-2a2926473303
- Vital Record Indicator: No
- Record Category Identifier: 2012-0000001111
- Review Period: None

Disposition schedule
General
Disposition Authority: UC Records Retention and Disposition Schedule, CCRC Internal
Disposition Instructions: Retain for 5 years, then transfer to Meritt for 5 years, then transfer to University Archives.
Applied to: Record
Unpublished Updates: No

Disposition Steps
1. Cut off immediately
2. Retain immediately
3. Transfer after 5 year(s)
4. Transfer after 10 year(s)

Actions
- Edit Metadata
- Manage Permissions
- Copy to...
- Move to...
- Delete
- View Audit Log

Share
This Page URL
https://hub-test-1.isi.berkeley.edu
Folder Disposition Rules

Edit Disposition Schedule

Steps

1. **Cutoff immediately**
2. **Retain immediately**
3. **Transfer after 5 year(s)**
   - Transfer to: **CDL Merritt Repository**
   - After a period of **5** Year from **Cut Off Date** or
   - When event occurs
     - Event
       - No longer needed
     - Event Completion
       - Manual
   - Which ever event is earlier

Step Description:

Transfer to Merritt after 5 years.

4. **Transfer after 10 year(s)**
Q&A, Contacts

• IST Research & Content Technologies
  • Patrick McGrath  patrickm@berkeley.edu

• University Archives
  • David Farrell  dfarrell@library.berkeley.edu
  • David De Lorenzoo  ddeloren@library.berkeley.edu
  • Josh Schneider  jschneid@library.berkeley.edu

http://hub.berkeley.edu/page/site/recordsman