We are living in exponential times.

There are about 540,000 words in the English language . . .

About 5 times as many as during Shakespeare’s time.

More than 3,000 new books are published . . .

Daily.

It’s estimated that a week’s worth of New York Times . . .
Contains more information than a person was likely to come across in a *lifetime* in the 18th century.

There are over 70 million unique visitors to Facebook each month.

The average MySpace page is visited 30 times a day.

There are over 6 billion searches performed on Google each month.

The number of text messages sent and received every day exceeds the population of the planet.

During the 300,000 years of human history prior to 1998, we created 12 exabytes (12 x 10^{18}) of information.

That’s roughly equal to 6,000 times the content of all U.S. Academic research libraries combined.

The next 12 exabytes of information was created in... 2.5 years.
In 2006 we created 150 exabytes.

In 2011 we expect to create nearly 1,800 exabytes.

Put another way, in the year 1930, information content was doubling approximately every 30 years.

By 2011 it will be doubling...

Every 11 hours.

In 2001 we created 2GB of information for every person on earth.

In a lifetime, a person can only read and process 4GB of information.

The amount of new technical information is doubling every 2 years.

It’s predicted to double every 72 hours by 2010.
That means for a student starting a four-year technical or college degree...

Half of what they learn in their first year of study will be outdated by their third year of study.

Third generation fiber optics that have been separately tested by NEC and Alcatel...

Third generation fiber optics that have been separately tested by NEC and Alcatel...

Pushes 10 trillion bits per second down one strand of fiber.

That’s 1,900 CDs or 150 million simultaneous phone calls every second.

Capacity is currently tripling about every 6 months and is expected to do so for at least the next 20 years.

Predictions are that e-paper could become cheaper than real paper.

That’s 2.3 million every second.

The volume of e-mails has exploded in recent years with over 200 billion now being sent daily around the globe.
177 million laptops will be shipped worldwide in 2009.

The “One Laptop Per Child” project’s goal is to put an inexpensive computer into the hands of every child.

Predictions are that by 2013 a supercomputer may be built that exceeds the computation capability of the Human Brain . . .

By 2023, a $1,000 computer could exceed the capabilities of the Human Brain . . .

Today’s Second Grader will be just 23 years old and beginning her (first) career . . .

And while technical predictions farther out than about 15 years are hard to do...

Predictions are that by 2049 a $1,000 computer may exceed the computational capabilities of the human race.

What does it all mean?

Shift Happens.
**Problem Statement:**

- Dartmouth College is failing to effectively manage the vast majority of the institution's electronic information.
- We are currently failing to manage our digital content as records.
- Are we unusual in this? No.

**Immediate results of this failure include...**

- Legal liability
- Lack of administrative control
- Inefficiency, inaccuracy, lost productivity
- Uncontrolled growth

**And some really bad long-term headaches!**

- Loss of information of enduring value (archival data)
- A crowded "virtual attic"
- Our names becoming swear words to our successors!

**Session Outline for Today**

- Review: Structured and Unstructured Data
- Document Management Technology
  - Overview
  - Features
    - Record Declaration
    - Record Classification
    - Implementation Issues
- What Dartmouth is doing...
- Some practical advice (while you're waiting...)
- Questions and Discussion

**Introduction**

- Your Name
- Department
- Job
- What makes you interested in digital record issues?

**The Records Management Training Cycle**

- Part 1: Managing Records in an Office Environment (June 2)
- Part 2: Using Records Management Services (June 9)
- Part 3: Vital Records Identification Protection, and Disaster Recovery (June 16)
- Part 4: Introduction to Digital Records (June 23)
- Part 5: Digital Document Management (Today!)
Structured vs. Unstructured Content

**Structured Content**
- Large, mainframe databases
- Institution or workgroup wide data repositories
- Examples: BANNER, HRMS, OASIS, ADVANCE
- Verdict: Mostly under control

**Unstructured Content**
- These are "documents"...
- In YOUR computer!
- "Documents are the containers in which information is presented for human processing"
- We have no single view of file organization
- 7.5 billion per year (2004)
- Here is where our challenge lies!

Structured vs. Unstructured Content

- Why are we concerned about "unstructured content"?
- The records exist outside of any centralized control of the institution
- They aren’t in central repositories
- They aren’t in any well-defined or consistent cataloging scheme
- It’s almost impossible to apply retention and disposition controls
- They are INVISIBLE to the institution
- At least with paper, it’s visible, you can see it, and you have to do SOMETHING with it!

The Continuum of Structured vs. Unstructured Content

Structured vs. Unstructured Content

Two reasons why getting control of our documents matters...

- 95% of business decisions are based on a document
- Paper transactions are 24x more expensive to process than digital transactions

Who has responsibility? (The Historical Perspective)

- Early Implementation
- Technology Infrastructure (Network, Software, Terminals, Supported Desktops)

Who has responsibility? (The Historical Perspective)

- Today
- Personal Computers

IT has not been designed with RM in mind!

- E-mail systems make purging easy, and declaring and saving a record difficult
- Web forms separate user context from information content
- E-mail and digital documents can be altered with ease
- Sharing documents proliferates more uncontrolled copies
- Hardware and software obsolescence foils preservation
- Digital media have short life spans
- Voice mail systems make it difficult to retain messages, even if they are contractually significant
Enterprise Content Management (ECM) Technology

Overview

Enterprise Content Management and Digital Records Management

The Big Picture...

CAPTURE
Document Ingestion and Creation

SELECTING and implementing technology to keep the information safely stored.

STORY

DELIVER
Getting the information where it needs to be!

PRESERVE

What are your actions for long-term archival and storage of your company’s essential content?

Archival Types

Bringing it all together...

The Big Picture...
Two ways data gets in to your repository:
1. Born Digital
   - Documents you create and declare
   - Received digital documents from outside
   - E-mail (including attachments)
   - Automatically generated digital reports and other content

Two ways data gets in to your repository:
2. Document Imaging
   - Converting Paper to Digital!

We now have content...
- ...so what can we DO with it???
  - Organize
  - Categorize
  - Search
  - View
  - Index
  - Annotate
  - Retain
  - Dispose
  - Define access rights
  - Create "virtual files"
  - We bring it under control
  - We create STRUCTURED data

Merging Repositories
Digital Document Repository
- Structured Data
  - Born-Digital Documents
  - Paper Files
  - Application Integration

Enterprise Content Management (ECM) Technology
Some Sample Features

Three Application Architectures
- Thin Client
  - Web Browser
- Thick Client
  - Desktop
  - Application

Organize and Categorize
METADATA!!!

Document Type Classification: A Digital File Plan!

WARNING!!!
- Although the technology I am about to describe is widely available and very stable, it is not yet available at Dartmouth.
- Successful ECM implementations require institutional commitment.
- If what I describe here looks useful, it’s up to us to advocate to get it!
Find what you need through custom searches

Browse your documents... without having to open them!

...or open them using native application support

Markup and manipulate your documents

Optical Character Recognition and Full Text Indexing

Application Integration

Structured and Unstructured Data come together!
From Three Repositories…
To One!

INTEGRATED ENTERPRISE CONTENT MANAGEMENT

Enterprise Content Management (ECM) Technology

1. The Record Declaration

Enterprise Content Management is about two things...
1. Declaration
2. Classification

Enterprise Content Management (ECM) Technology

Take a Break!

* Be back in five minutes...

The “Record Declaration” Defined

- “Freezes” the record (It is no longer editable)
- Distinguishes a record from a non-record or a “working paper”
- Record should move to a repository
- The maintenance of the information becomes a departmental (and institutional!) responsibility

Declaration - From a Digital Perspective...

- Put a document under e-Records Management Control
- The document is no longer editable and can no longer be deleted outside of the retention schedule
- Analogous to deciding whether or not to file a paper document

Declaration Options

<table>
<thead>
<tr>
<th>Automatic Declaration</th>
<th>Manual Declaration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pros</td>
<td>Cons</td>
</tr>
<tr>
<td>No reliance on user</td>
<td>Difficult to implement</td>
</tr>
<tr>
<td>No missed documents</td>
<td>Could miss valuable records</td>
</tr>
<tr>
<td>Easy to implement</td>
<td>You will never know which records you missed</td>
</tr>
</tbody>
</table>

Pro

Cons

Declaration Options

Wess Jolley, CRM, Records Manager
Declaration – Automatic
- Declaration triggered by some system event such as:
  - A step in a workflow
  - Intercepting a sent e-mail
  - Placement in a monitored folder
  - Triggering date of creation (or period afterwards)
  - Anything else you can think of!

Declaration – Automatic E-mail Example
- Subject Lines
- Recipient List
- Complex Logic
- Pros
  - Easy to implement
- Cons
  - Only works for SENT e-mail
  - Semi-Automatic: Force User to chose every time they click “send” (optional bail out?)

Declaration – Automatic By Folders Example
- Monitored Folders
- Preconfigured Folders and Profiles
- Declared upon ingestion
- Pros
  - No reliance on user
  - EASY to implement
- Cons
  - All possible folders must be set up in advance
  - It equates to manual classification

Enterprise Content Management (ECM) Technology
2. Classification

Classification Defined
- The application of a predetermined file plan and associated retention schedule to declared documents in a document management system
- Putting documents into record series
- “Place a document in a bucket”
- Automatic or Manual
- Analogous to deciding where to file a paper document

Classifying a Document

Classification Options
- Automatic Classification
  - Pros
    - No reliance on user
    - Accuracy and consistency
  - Cons
    - Not as easy to implement

- Manual Classification
  - Pros
    - Easy to implement
    - Works well when done right
  - Cons
    - Difficult to implement
    - Not as easy to implement

- Allows for user interpretation
Classify – Automatic

- Metadata or content based
- Pros
  - Easy to implement
  - Simple, instant
- Cons
  - Depends on Metadata quality and consistency (abbreviations, spelling, etc.)
- Good metadata is tightly controlled (data entry, automatic generation)
- Bad metadata is uncontrolled (such as e-mail subject lines)

Classify – Manual

- Browse File Plan
  - Navigate (browse) anywhere in file plan
  - Presents greatest choice, greatest effort
- Quick List
  - Pre-configure a list of selections for a user/group
  - Add some logic
  - Minimal Choice, low effort
  - The list has to be right...

Classification – Manual By Folders Example

The “Holy Grail” of Classification: Content Based Auto Classify

- Software that “reads” a document, reduces to core concept
- Compares document concept to File Plan
- Finds the best match
- Builds a taxonomy on the target subject
- Self-learning and dynamic
- It doesn’t exist—yet (accuracy still around 85%)

A Fantasy (But Realistic) Example of Declare and Classify

- Declare a document
  - Voluntary “Declare” Button
- Classify
  - Manual
    - Browse the file plan
    - Select from designated pick list
  - Automatic
    - Email containing a specific word in the subject line

Ten Steps to a Successful ECM Implementation

1. Build supporting organizational structures
2. Build institutional awareness
3. Establish supporting policies
4. Enshrine the new policies
5. Build/ Strengthen RM Foundation
6. Develop implementation strategy/plan
7. Map business processes
8. Implement RM Technology
9. Conduct the initial pilot
10. Enterprise roll-out

Enterprise Content Management (ECM) Technology

Implementation Issues

Effort vs. Reward for End Users: The Cultural Reality

- Effort
  - Declare the document
  - Classify the document
- Reward
  - NONE!
- Resistance
  - This is my document
  - It is too much work
- The Five Second Rule
  - “Filing a document into a records repository is an unnatural act” (R. Medina, 2000).
Planning a declaration strategy
- When do we *want* to declare?
- When is the best *opportunity* to declare?
- When is a user most likely to *agree* with a declaration?

Planning a Classification Strategy
- Do we have reliable metadata?
- How/when can we grab the metadata?
- What kind of rules can we construct?
- How reliable will those rules be?
- How vulnerable are we to bad metadata?

Planning an Overall Strategy
- Are we ready?
- What is our cultural reality?
- What is our user attitude? Should we even consider manual steps?
- Process Sensitivity. What will be the impact on the work process?
- Criticality. How important are these records?

What Dartmouth is Doing …

The Digital Document and Records Management Task Force
- Membership
- Methods
- Goals
- Process

Survey Results…

Question 1
Does your department consider digital documents to be records?

Question 2
Is there a plan in place to manage digital records?

Question 3
What’s your current policy on paper and digital records?

Question 5
Do you consistently apply your records retention schedule?
Digital Document Management

Wess Jolley, CRM, Records Manager

June 30, 2009

Question 6
Digital documents: stored on a server or on individual workstations?

Q6 results

<table>
<thead>
<tr>
<th>Stored on Server</th>
<th>Stored on Workstations</th>
<th>No Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>47%</td>
<td>36%</td>
<td>17%</td>
</tr>
</tbody>
</table>

“Bonus” question
Would you be interested in learning more about digital records management?

“Bonus” Q results

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>78%</td>
<td>22%</td>
</tr>
</tbody>
</table>

Findings...
- Every area of the College is facing an ongoing transition to digital record keeping technology
- Workflow, collaboration, and data sharing are important aspects of records management, and the resources needed to support them are not emerging.
- Digital documents and records are at risk and are not being managed properly due to their dispersal across individual workstations.

Findings...
- Existing digital systems provide no facility for documents to comply with existing record retention policies. This results in the potential for data loss, and/or legal liability for the institution.
- Digital record keeping demands are outpacing resources and technology.
- Administrative and academic departments are under pressure to upgrade their systems to more sophisticated and efficient digital technology.

Institutional Needs
- Data sharing
- Security
- Collaboration
- Retrieval
- Life Cycle Management
- Preservation
- Workflow
- Versioning
- Data Storage
- Leadership
- Others?
D2I: Dartmouth’s Digital Data Initiative

- The Duke Collaboration
- D2I: Digital Document Initiative
- The Future?

What we want to avoid...

- Ad hoc implementation
- Huge variety in software applications and strategies
- Large maintenance burdens
- Multiple small site licenses
- No ability to share, combine, and grant access to repositories

What we want to achieve...

- A technological infrastructure to which the institution is committed
- A strong support network
- Ability to achieve quick deployment
- An institutional strategy!

Some Practical Advice

...while you’re waiting and planning

- Start thinking NOW about the future
- Dream big!
  > Systems that incorporate both paper and digital documents into one repository
  > Systems that combine structured and unstructured data
- Get your Plane on the Runway!

Some practical advice

- Start building support for change
  > Staff, reporting structure, top administration. TALK about it!
  > Become a Digital Records Evangelist!
- Begin getting the infrastructure in place
  > Document repository servers
  > Define security needs
- If you’re really chomping at the bit, talk to me!

PDF Handouts

http://www.dartmouth.edu/~library/recmgmt/forms.html

What We Have Learned...

- Review: Structured and Unstructured Data
- Document Management Technology
  > Overview
  > Features
  > Record Declaration
  > Record Classification
  > Implementation Issues
- What Dartmouth is doing...
- Some practical advice (while you’re waiting...)

Questions and Discussion
Graduation!

Theresa Ciambru
Lisa M. Daoust

Congratulations
to the Class of 2008!!!