Did You Know . . . 

We are living in \textit{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Half of what they learn in their first year of study will be outdated by their third year of study.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...

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.

The volume of e-mails has exploded in recent years with over 200 billion now being sent daily around the globe.

That's 2.3 million every second.
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.
Now you know . . .

Original presentation created by Karl Fisch
http://thefischbowl.blogspot.com/2006/08/did-you-know.html
Original source available at:
http://www.lps.k12.co.us/schools/arapahoe/fisch/
didyouknow/sourcesfordidyouknow.pdf
Version 2.0 of the presentation available at:
Additional material added by Wess Jolley, CRM
wess.jolley@dartmouth.edu
Sources available upon request

How much data are we creating?
- A lot!
- The world has never experienced such massive change and growth in how it does business, and how it manages information, as has been seen in the last 20 years.
- And this rate of growth is accelerating!

InformationWeek
The Digital Universe Created 101 Exabytes Of Data Last Year
The amount of digital media created last year is equal to 10.1 zettabytes. The total amount of all human knowledge, including all books ever written, plus all photos, music, and movies, is about 7 zettabytes.

Introductions
- Your Name
- Department
- Job
- What makes you interested in digital records 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 (Today!)
- Part 5: Digital Document Management (June 30)

Session Outline for Today
- Overview: What is a Record?
  - Part 2: Overview of Some Key Emerging Technologies and Issues
  - Part 3: Some Practical Advice
  - Conclusions and Questions

Records Management Manual
http://www.dartmouth.edu/~library/recmgmt/

“What is a Record?”

Wess Jolley, CRM, Records Manager
“Any recorded information generated in the course of conducting business, which must be maintained to meet the fiscal, legal, historical or administrative needs of the organization.”

Records should be declared and classified as part of a formal process or workflow. Records exist regardless of their format.

**RECORDS:**

   - Let’s examine some analogies...
     - How is a computer like a...
       - Typewriter?
         - A record creation engine
       - A Filing Cabinet?
         - A record storage device
     - The Death of the File Room!
       - The easy way out:
         - ONLY a record creation engine
         - NOT a records storage device
         - Digital files are NOT records
         - Digital information only BECOMES a record when it is printed out
         - (Printing becomes the record declaration)

   - Are Digital Files Really “Records”?
     - The realistic approach:
       - Acknowledge that computers are both:
         - A record creation engine
         - A record storage device
   - We must acknowledge that we’re long past the era when all transactions produced a paper record
   - Do the courts accept digital records?
     - Yes! The courts have decided that digital records are REAL records
   - E-mail is included!
     - Proper maintenance and destruction of e-mail records is vital—and challenging!
     - E-mail can and is being used as evidence!

**E-mail is included!**

Proper maintenance and destruction of e-mail records is vital. It can and is being used as evidence!
This is the Text of the E-mail that Destroyed Arthur Andersen, A Multi-Million Dollar Company

Mike-

It might be useful to remind the engagement team of our documentation and retention policy. It will be helpful to make sure we have complied with the policy. Let me know if you have any questions.

--Nancy

As quoted in "Information Nation", Randolph Kahn and Barclay T. Blair, pg. 51

The New Frontier: Digital Discovery

What is "discovery"?

If it exists, it is fair game for the lawyers and the courts!

IMPORTANT: It may exist, even if you think it doesn’t!

"Trashing" doesn’t remove the data

Sharp technicians can usually recover it all

If the stakes are big enough, they’ll get it

For example, in Murphy Oil USA v. Fluor Daniel, Inc, the court heard a dispute in which Murphy Oil wanted Fluor Daniel to go through nearly 20 million pages of email records to see if any of those records related to the case. The reason there were so many pages of e-mail to search through is that Fluor had apparently not been following its own policy, as the court noted:

"Fluor’s email retention policy provided that backup tapes were recycled after 45 days. If Fluor had followed this policy, the email issue would be moot. Fluor does not explain why, but it maintained its backup tapes for the entire 14-month period."

Fluor estimated that the cost of providing relevant documents from the 20 million pages of email and attachments would be in excess of $6 million, and would take six months — far more than the cost would have been if they had followed their own policy.

--Kahn, Randolph and Blair, Barclay T.; Information Management: Seven Keys to Information Compliance. 2004

Structured vs. Unstructured Content

Structured Content

- Large, mainframe databases
- Institution or workgroup wide data repositories

Examples: FAS, BANNER, HRMS, Oracle Financials

Verdict: Mostly under control

Structured vs. Unstructured Content

Unstructured Content

- These are "Documents"…
- In YOUR computer!

"Documents are the containers in which information is presented for human processing"

- 85% of all digital records in any organization
- 7.5 billion per year (2004)

Here is where our challenge lies!

For more on Digital Document Management, come back next week!

Compare these statistics:

- Over two-thirds of all information generated today is in digital format ("Born Digital"), 1
- Paper filing cabinet sales increase by 18% per year. 2

What happened to the "paperless office?"

We are not there for two reasons...

Reason 1:

We have not committed ourselves to managing digital documents AS RECORDS

- Traditionally, when we have sensed a "record value" for a digital document, what have we done?
- We’ve printed it out!

- We are more confident with managing the life-cycle of a document through paper:
  - Inactive filing
  - Retention Scheduling
  - Disposition
  - Etc.
Reason 1:
We have not committed ourselves to managing digital documents as records.

- But this complicates matters, because we now have two copies: paper and digital.
- Legal opinion tells us that both are records, in the eyes of the courts.
- We can be saved or condemned by either one.
- More importantly: if sued, we have to produce both!

Reason 2:
We still use paper to communicate between incompatible digital systems.

- An Accounts Payable Example

Paper System (1975)

Digital System (Soon!)

Change is Here!
- Legal Changes
- Technology Changes
- Human Changes
- Inevitability

Take a Break!

Part 2:
Overview of Key Technologies

Centralized Document Repository Servers
- Distributed Document Storage

Ivan

Allen

Sylvia

Wess Jolley, CRM, Records Manager
Centralized Document Repository Servers
- Distributed Document Storage
- Centralized Document Repository

Document Management
- We have CONTENT... Now what do we DO with it??
  - Organize
  - Categorize
  - Search
  - View
  - Index
  - Organize
  - Retain
  - Define access rights
- We bring it under control
- We create structured data
- We'll examine this in detail next week

Document Imaging
- Unifying the repositories!
- Converting paper to digital!

Digital Workflow

Electronic Forms
- PDF
- Web

Encryption and Security

Web Content Management
- Web Pages can be Records!
  - They can be records themselves
  - They are often next-generation engines
- Content and Context
  - Web sites typically strip context from transactions
- Recordkeeping options
  - Snapshotting
  - Versioning
  - Archiving of HTML with transactions
- Content Management Systems

Preservation

Requirements for Digital Preservation
- Hardware
  - Readable
  - Retrievable
  - Understandable
- Content Management
  - Identifiable
  - Authentic
- File Format
  - Encapsulated
  - Reconstructable
**Introduction to Digital Records**

June 23, 2009

**Preservation: Storage Media and Formats**
- Two truths:
  - All storage formats (particularly magnetic media) are unstable!
  - File formats WILL change!
- Committing to migration
- Developing a migration schedule
- When is it really worth it?

**Our Goal:**
- Worrying about this shouldn’t be your problem!
- Dartmouth should make document management services a key and core component of our digital strategy
- Attend our “Digital Document Management” training

---

**Part 3: Some Practical Advice**

*What little there is...*

**Centralize, Centralize, Centralize!**
- Utilize a central document repository for your workgroup or department
- Implement the policy that a "record declaration" means it goes to the server
- All other documents are "working papers" to be disposed as soon as practical

**Be Deliberate About the “Record Declaration”**
- What is a “Record Declaration?”
- Caveat: This is meaningful to us, but not to the courts...
  - Therefore, dispose non-records and working papers ASAP!
- Use the analogies we already know about working with paper...
  - Use files and folders for organization
  - View your hard disks as a filing cabinet
  - View your trash as destruction

**Do Not Exceed Existing Retention Schedules**
- Utilize your existing schedules where you can
- Develop specialized retention schedules for unique digital content

**Categorize, Categorize, Categorize!**
- Use existing tools
  - Series
  - Existing filing systems
- Develop new tools
  - Naming Conventions
  - Hierarchies
  - Taxonomies
  - Keywords
  - Search terms
  - Indexes
  - Folders within folders
- Match existing file plans
  - Parallel systems
  - Parallel retention
  - Parallel disposition

**Document, Document, Document!**
- Use your existing Records Management Manual
- Explicitly state that the policies in it apply to digital content
- Note specialized digital retention schedules
- Document retention and disposition actions
  - You can utilize the forms from Session 1: [http://www.dartmouth.edu/~recmgmt/forms.html](http://www.dartmouth.edu/~recmgmt/forms.html)
In July, a hedge fund manager at San Francisco company Azure Capital Partners LP reportedly IMed his AOL buddy list with information about PeopleSoft Inc. and was later accused of undermining PeopleSoft's stock price. The manager either said to buddy list members that regulators were looking into accounting irregularities at a PeopleSoft subsidiary or asked if that were so, and also raised the possibility that PeopleSoft was being sued by a customer for $50 million. The IM turned into a rumor, PeopleSoft got their hands on the message and concluded that it was the cause of what turned out, over the course of a few days, to be a $1.7 billion drop in the company's market value.


**E-Mail Management**
- Recognize that e-mail, IM and chat are special challenges, with high liability
- Watch what you say!
- Be discriminating in the records declaration
- Delete regularly
- Don't save to disk

**Manage Duplication**
- Identify the “Record Copy”
- Dispose of duplicates and drafts
- Think twice before preserving incoming documents
  - Is this a record in the originating location?
  - Can that copy suffice?
  - Do you need a copy as well?

**Snapshot “Fluid Files”**
- What is a “Fluid File”? Databases, primarily
  - Small (personal or departmental)
  - Word or Excel
  - Any collection of information that is updated
- Determine snapshot schedule

**Perform Regular Disposition Rounds**
- Minimize your legal liability
- Don’t get caught with documents outside of your retention schedule!
- Strive for secure deletion
  - DOD 5220.22-M
- At least empty the Recycle Bin!
- Keep up with it
- Document your actions

**Protect Your Vital Records:**
**BACKUP!!!**
- Backup is essential!
- Methodologies
  - Tape
  - CD
  - Data Warehouse or Server
  - Net Backup or on your own
- Regularity
  - How much can you afford to lose?
- Off-Site

**PDF Forms**
http://www.dartmouth.edu/~library/recmgmt/forms.html
What We Have Learned...
- What is a Record?
- Digital Record Keeping: The Challenges of a Changing World
- Overview of Key Emerging Technologies and Issues
- Some Practical Advice

Questions and Discussion