16. Institutional / Enterprise Information Management

INFO 202 - 22 October 2008

Bob Glushko

Plan for Today's Lecture

Knowledge management

"Unstructured text" management

Content management

- AGU publishing case study

Records management

- Sarbanes-Oxley and other compliance mandates

Information Management and the Document Type Spectrum

Knowledge Management

Much collective knowledge is embodied in a firm's people, systems, management techniques, history of strategy and design decisions, customer relationships, and intellectual property like patents, copyrights, trademarks, brands, etc.

The goals of KM can be viewed as getting the tacit parts of this "intellectual capital" to be explicit

- Sharing solutions to customer problems
- Facilitating collaboration
- Locating people with relevant skills
- Managing unstructured content
- Providing greater access to existing information
- Improving traceability and justification for strategic (and controversial) decisions
- Recording the rationale for business process and information models

Additional Public Sector Goals for Knowledge Management

Improved efficiency in procurement to reduce costs to taxpayers

Improved traceability and justification for controversial decisions

Efficiently satisfy information requests

Knowledge Management Approaches

Many technologies have been used for KM -- Lotus Notes, Intranets, Wikis, Blogs...

But at best, knowledge management techniques can only capture knowledge that is codifiable and transferable, and not all knowledge is

And furthermore, employees have complex motivations for complying with or not complying with KM goals

"Enlightened" firms and management try to align personal and corporate goals for knowledge management through "assetization"

The Mandate for Managing "Unstructured Text"

UTAA Application	Textual Data Source
Business intelligence	Web, industry blogs, online databases
Customer relationship management	Customer feedback, help desk reports
Regulatory compliance	All internally generated electronic documents
Intellectual property management	Web, copyright and patent databases
Call support (help desk applications)	Call documentation, customer feedback, email, online manuals
Accounts payable/receivable analysis	Invoices, customer and vendor correspondence (used frequently with traditional structured data mining and analysis)
Legal department support	Legal databases, specific streams of organizational communications (such as customer communication, internal email)

"Once Internet search became available, people expected the same level of availability in their business lives"

Typical Functions / Processes in UT Applications

Document management

- Storage of original document in "native" format
- Conversion to "canonical" format

Analytic processing to extract words and concepts

- Indexing to enable search
- Concept / entity extraction using statistical "concept discovery" techniques or domain-specific dictionaries, thesauri, and ontologies
- (We will spend lectures #22-28 discussing these approaches)

Content Management

"Content Management" narrowly defined involves the management of semi-structured content in a logical repository, usually in a multi-user collaborative context

But "content management" necessarily involves authoring and delivery or there would be nothing to manage or no purpose in managing it

"Flavors" of Content Management

Document management

Web content management

Digital asset management

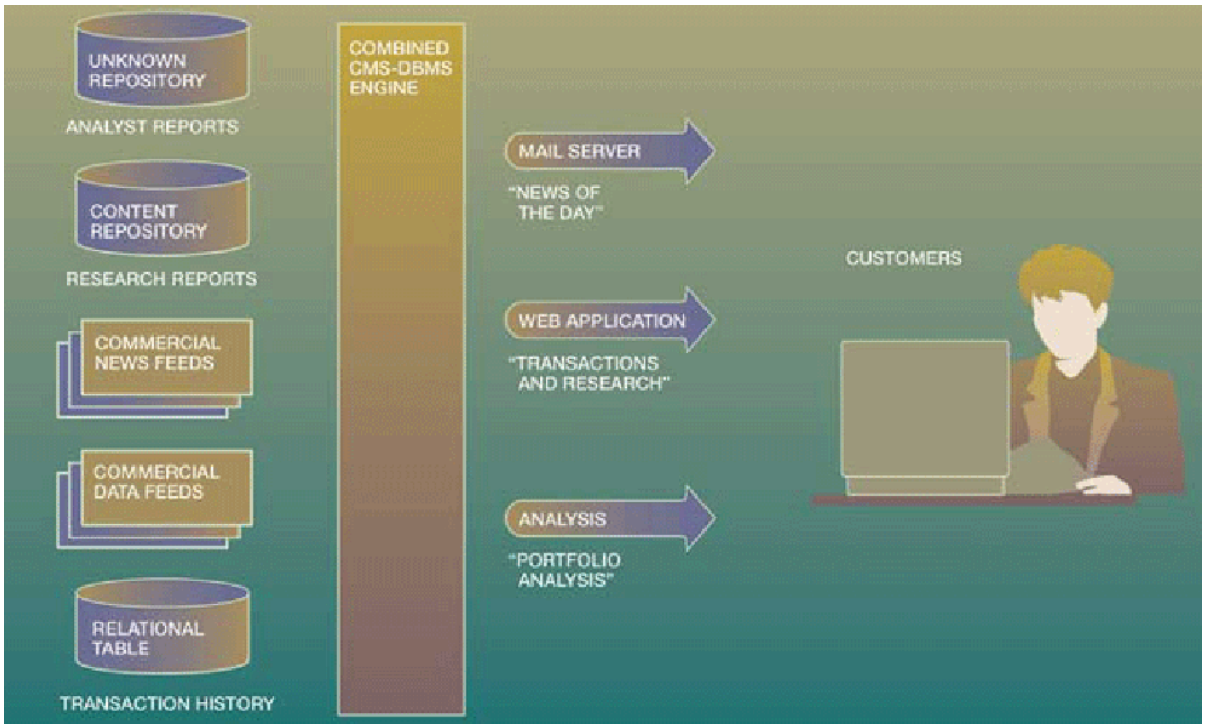
E-mail management

Records management

Report management

Collaboration tools (Notes, Wikis)

Note: "Content" and "Data" Management Are Often Intertwined



Who Needs Content Management?

Is there a high volume of content?

Is it created by many authors, both alone and in collaboration?

Are there multiple instances of the same or closely related document types?

Are multiple document types or formats required for different contexts, users, or devices?

Does the content have a long useful life?

Is the production, management and use of the content governed by formal processes or regulation?

Content Authoring

Authoring can be broadly defined as creating reusable "information assets" from different sources

Reusable information sometimes means XML, but more generally means information objects with metadata

Reusable information assets can be created by adding structure and metadata to existing information

Non-text information assets can be described using XML text metadata

Content Management (narrow sense)

Reliable storage and retrieval of components, documents, schemas, transforms, stylesheets...

Componentizing a document by separating it into its constituent elements using user-defined names as boundaries

Risk management functions like backup and archiving

Component Granularity

What level of granularity is desired / required / achievable?

- Document level granularity
- Module level granularity
- Content unit level granularity
- Word level granularity

Content Delivery

Content delivery is fundamental to the business models of news services, publishers, sellers, distributors, etc.

Content delivery usually begins when some set of components is retrieved from the repository and assembled to meet some specific requirement

Assembly may involve both the assembly of a document type model and then the assembly of an instance that conforms to it

The retrieved or assembled instance may need to be transformed to conform to another model

A Single-Source Strategy

Single-source is a popular slogan in content management that has both informal and rigorous definitions

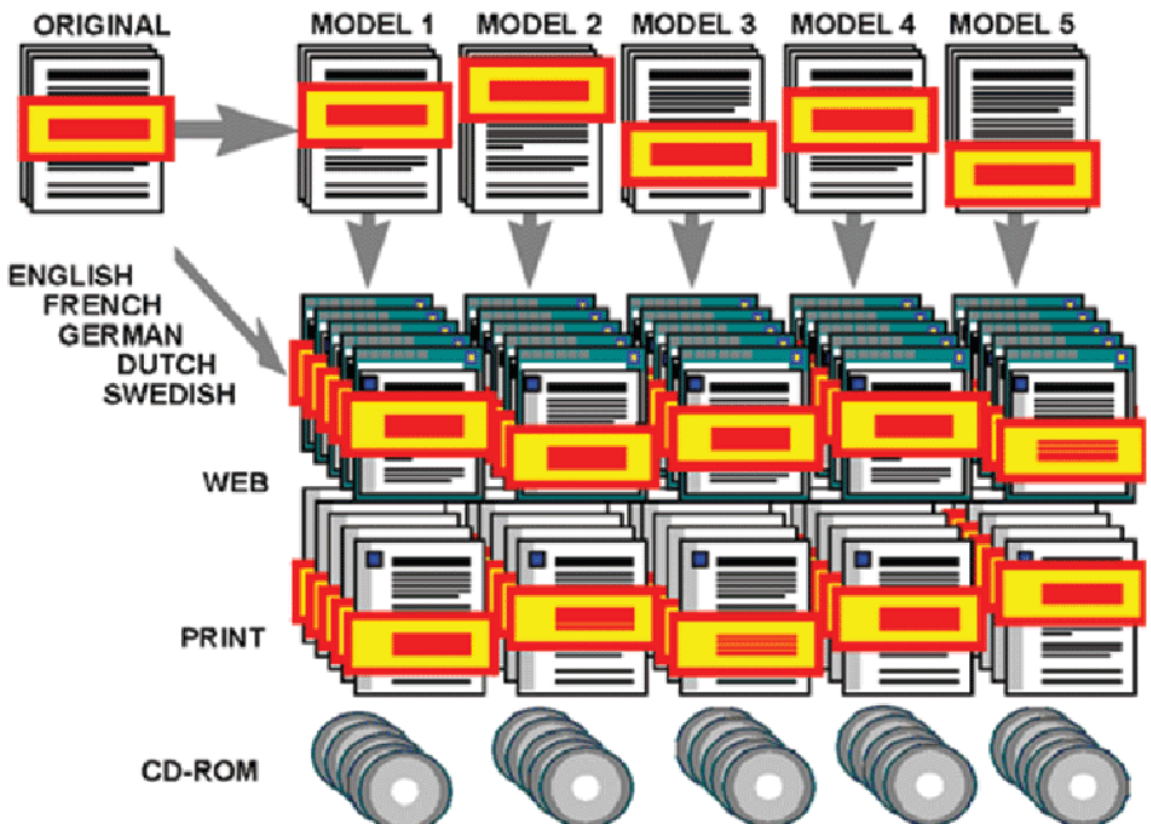
Informal:

- Write once, reuse many times
- Revise once, update everywhere
- Transform many times for delivery

Rigorous:

- Enforce normalization techniques to prevent anomalies with duplicate content
- Use transformations to convert content from one structure or context to another, storing the transformations rather than their results

When You Should Single Source



Implications of Single-Sourcing

Documents are much more stable because content changes are controlled

Document consistency and quality should improve because of inherent automation of change propagation and assembly

The total number of content components might increase significantly because they will be smaller

Policies and practices must be developed and enforced, but not all of this enforcement can be done by automated means

The organization of people and tasks may need to be changed to make single-sourcing work

AGU Case Study

How American Geophysical Union redesigned its publishing processes and technology

- Substantially increased productivity in producing existing publications
- Enabled many new kinds of publications

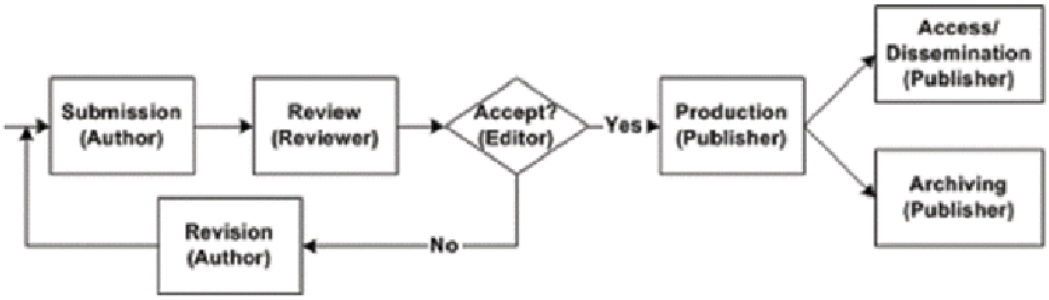
Well written technical case study with fascinating business and organizational "texture"

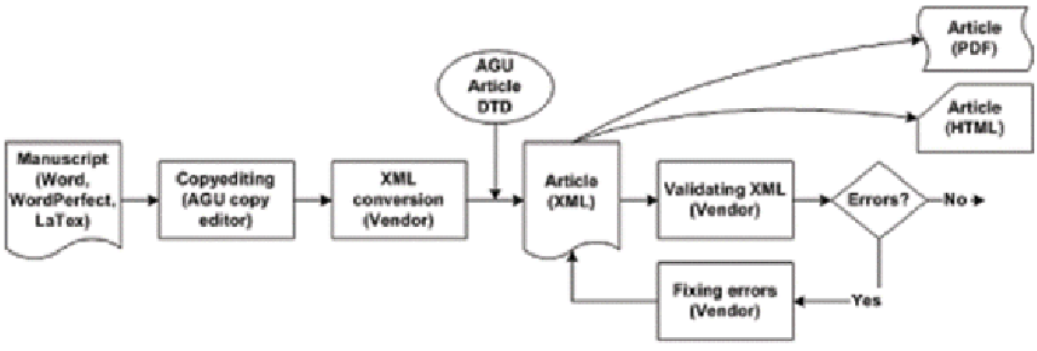
AGU Project Goals

Develop an information creation, management, and delivery system that

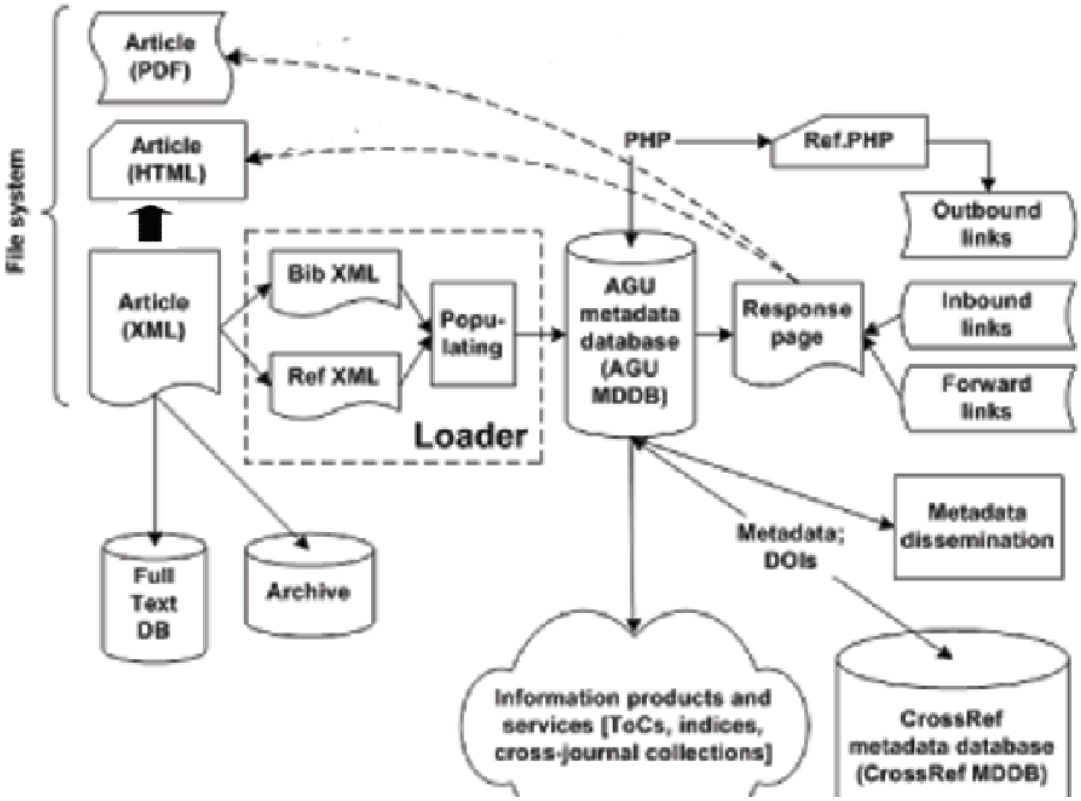
- Is a single logical repository to eliminate duplicate authoring and distribution
- Makes it easy to update document components and deliver them as needed with little human intervention
- Reuses common information across different product lines

AGU Authoring -- Before and After



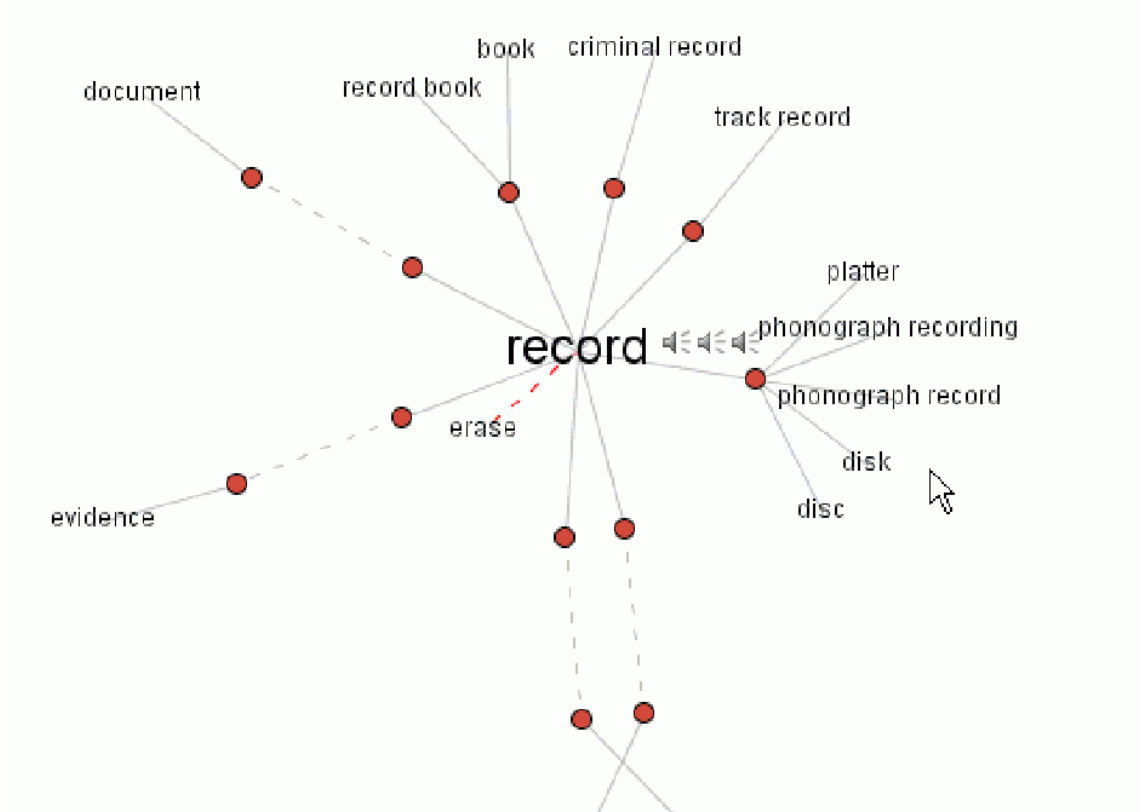


AGU Publishing System

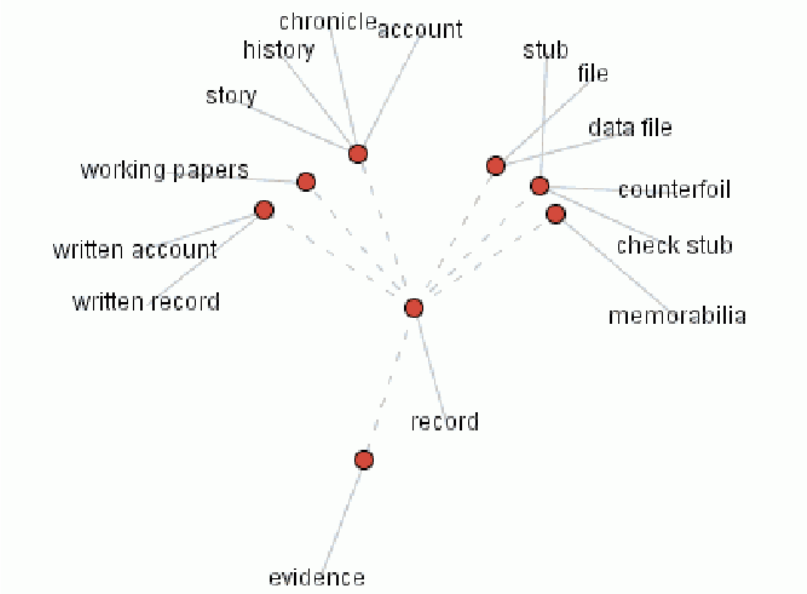


The Computerization and Automation Paradox

A Record is a Type of Document



Records are "Permanent" Evidence of Events



Records

Records may be created on any physical media including:

- Paper
- Film (microfilm, photographic film, x-ray)
- Disk (optical, magnetic, video, audio)
- Tape (magnetic, video, audio)

The method of recording may be manual, mechanical, photographic, or a combination of these technologies

Records Management

"Content and process management are inextricably linked via records management" (Barbero and Douglas)

When does content become a business record?

Retention requirements

Non-retention requirements

Purging requirements and purging authority

Recordkeeping Problem Areas [1]

Documentation of policy and decision making accomplished orally or electronically

- "... require personnel at all levels to document conversations and meetings dealing with significant program business by preparing a detailed and signed memorandum or form identifying the participants and summarizing the conversation or meeting"
- "If records...do not show the complete names of senders, addresses, and the date of transmission,users should take reasonable steps to preserve the mail envelope, distribution lists..."

Recordkeeping Problem Areas [2]

Contractor records

- "Unless contract provisions explicitly define the documentation to be provided to the agency, contractors are likely to create needed documentation as private property"
- EXAMPLE: Until late 1980s when the government acquired "software" contractors would deliver only the object code and no documentation

Didn't Follow Federal Recordkeeping Rules

Oliver North's Recordkeeping Mistake

Oliver North used the White House email system to conduct one of the most scandalous activities ever carried out by the US government.

In 1986 several members of the Reagan Administration sold weapons to Iran, an avowed enemy, and used the proceeds to fund the Contras, an anti-communist guerrilla organization in Nicaragua

To conceal his involvement North and his secretary Fawn Hall shredded all pertinent papers and deleted all relevant e-mail.

North didn't realize that e-mail was backed up, and the e-mail was used as evidence against North

Recordkeeping Problem Areas [3]

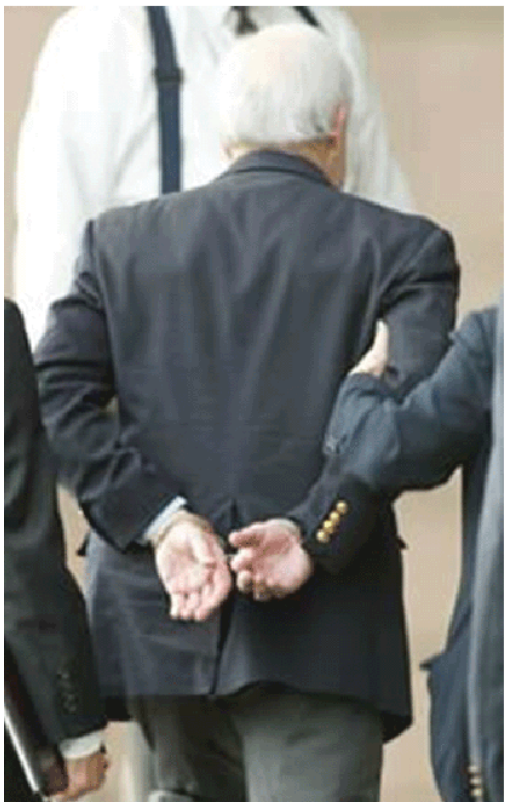
Personal papers and files

Documentation of formal meetings

- EXAMPLE: Cheney's "Energy Task Force" (http://www.judicialwatch.org/5309.shtml) did not have to disclose names of participants

Drafts and working files

Ken Lay Does the Perp Walk



Sarbanes-Oxley and Information Management [1]

The Sarbanes-Oxley Act of 2002 was enacted to curb corrupt business activities and fraudulent accounting practices like those of Enron and WorldCom.

SOX (aka Sarbox) requires firms to implement adequate internal control structures and procedures and attest to their effectiveness.

SOX requires sufficient auditing and traceability to relate the IT systems that carry out internal controls and the financial reporting process to the firm's financial statements

Complaints about cost of compliance made some weakening of SOX provisions likely, but 2008's financial meltdown is likely instead to increase regulatory and compliance requirements

"Material" Information

Sarbanes-Oxley and Information Management [2]

SOX also requires that firms disclose "material" information about their operations and financial situation in a timely and predictable manner ("trip wires") that trigger disclosure

So SOX is causing is causing increased spending in document and records management, security, business process management and document engineering as companies define, document, and automate the processes that are needed to run the company while enabling auditing and timely reporting

Standardization underway to develop an "Extensible Business Reporting Language
(xbrl.org) and standard models for the auditing document types and their interrelationships

EXAMPLE standard timesheet instance (http://www.gl.iphix.net/)

The "Reasonable Man" Standard

Many firms, especially small ones, complain that compliance costs impose excessive burdens

Some people argue that a firm should treat these costs as strategic investments in more effective business processes

But others argue that full and failure-proof compliance isn't achievable, advocating a "reasonable" or "managed risk" approach that involves data encryption, access controls, process documentation for monitoring and tracking, and employees trained to make use of them

The Stolen Berkeley Laptop

Laptop theft at UC Berkeley

On 11 March 2005 someone stole a laptop from an office in Sproul Hall that contained personal information about 98,369 alumni, graduate students and past applicants

"For several years University of California systemwide policy and UC Berkeley campus policy have required that restricted information stored on portable equipment be protected to safeguard the data if the equipment is lost or stolen. Since fall 2004, the UC systemwide policy has required encryption of such portable data, and campus units are in the process of moving toward full compliance with this new policy."

The Stolen Berkeley Laptop: Lessons?

"Our challenge is not that we lack policies governing computer security and the safeguarding of sensitive information. Our policies are clear, and during the last fifteen months we have strengthened them. Our challenge is enforcing these policies, and specifically, rectifying the lack of clear lines of accountability, both personal and departmental."

Some Summary Thoughts About Knowledge/Content/Data Management

Information technologies can solve many of the problems of knowledge, content and data management but can also cause them

Information technology has radically transformed the nature of business so that every enterprise of significant size, regardless of industry, must view these challenges as critical

Enterprise concerns are driven by internal goals like efficiency and core competency and also shaped by external factors like competition and compliance requirements

These concerns are moving up the company hierarchy; many firms have CIOs and increasing numbers have CPOs and CKOs

Midterm: Monday 27 October

Choice of short answer questions

Open book, open note, open "study resources"

"Cut and paste" strongly deprecated; quality of answers more important than quantity

If worried about time... write "bullet points" first and then integrate into coherently argued prose