16. Institutional / Enterprise Information Management [1]

INFO 202 - 22 October 2007

Bob Glushko
Plan for Today's Lecture

Content management
  - AGU publishing case study

Records management
  - Sarbanes-Oxley and other compliance mandates

Enterprise Data Management via MRP and ERP
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 (in the 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 (the fix)
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
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.
"Material" Information

OUR SHAREHOLDERS ARE SUING US FOR MISLEADING THEM ABOUT OUR FINANCIAL PROBLEMS.

SINCE WHEN IS IT ILLEGAL TO SHAFT INNOCENT PEOPLE FOR PERSONAL GAIN?

DON'T PUT THAT IN THE MINUTES.

I'LL SEE WHAT I CAN DO.
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 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."
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.
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."
"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."
Enterprise Data Management via MRP and ERP

Business processes can span multiple departments, business applications, or even multiple firms.

Internal to a firm, these "silos" or "stovepipes" may have been created over time and not have been designed to share information.

Each of these systems has a specific purpose and a data model customized for that purpose - so these models may be incomplete or incompatible.

Connecting different applications to share information has long been a substantial portion of the IT activities in many companies.

One alternative "solution" to this integration problem is to replace the separate applications with a single enterprise-wide application that contains all the functionality needed by the company.
MRP – Materials Requirements (or Resource) Planning

MRP was the big business buzzword in the 1980s

MRP systems plan production, procurement, and distribution for an enterprise

For each of the products a company manufactures there is a "recipe" or "parts list" or "bill of materials" that lists the parts or components that go into it (and how many of each).

"Shopping lists" for each production cycle then sent to suppliers

```
BILL OF MATERIALS  X SALES FORECAST = PARTS WE NEED
INVENTORY ON HAND = PARTS WE ALREADY HAVE

PARTS WE NEED - PARTS WE ALREADY HAVE = PARTS "SHOPPING LIST"
```
ERP – Enterprise Resource Planning

ERP was the big business buzzword in the early 1990s

ERP is a natural evolution of MRP that connects it to other key functions of an enterprise – it is the information "backbone" or "nervous system" of a big firm

ERP interconnects internal systems for manufacturing control, production planning, inventory, and procurement (scope of MRP) with accounting, finance, and personnel

ERP systems provide step-by-step guidance for the processes that aren't automated

ERP can enable substantial efficiencies and collaboration because each function can more readily see how it affects and is affected by the other
Benefits of ERP

The end-to-end integration of ERP ties purchasing decisions to the organization(s) that makes them and imposes financial controls on all the processes.

ERP eliminates waste in the production and distribution of goods and reduces excess and obsolete inventories.

ERP can be used as a working model or simulation of the firm.
Implementing ERP

But implementing ERP is hard and many ERP implementations fail (completely or partly)

Some implementation difficulty is architectural – ERP software evolved in the 1980s from single programs running on mainframes to distributed client-server architectures, but the various modules remained very tightly coupled to the core system

But most difficulty is intrinsic to the problems ERP tries to solve – automating and enforcing the business rules of an enterprise by integrating a company's legacy computing and computer-controlled operations

This usually involves changing business processes and information models ... which effects how people do their jobs ... and they resist
Readings for INFO Lecture #17

Larry Downes, Chapter 4, "The Information Supply Chain," The Strategy Machine


Judith Hurwitz et al, Chapter 13, "Where's the Data?, Service Oriented Architecture for Dummies, 2006

Anant Jhingran, Enterprise Information Mashups: Integrating Information, Simply. VLDB '06.