White Paper

Enterprise Content Management Systems

January 14, 2008

Intel Confidential

Project Name: IT Storefront CMS Evaluation

Project Sponsor:	Donn Kanagaki	**Phone #:**	916-356-4029
Client Organization:	IT Marketing	**Billing BGL #:**	55748
Program Manager:	Pier Gittens	**Phone #:**	916-356-6730
Project Manager:	Noel Gifford	**Phone #:**	503-264-0714
Content developed and reviewed by:	Jim C. Cross	916-356-4586	BAS
	Michael Talman	916-356-4011	BAS
	Gary Behan	480-554-2415	BAS
	Erica Tierney		BAS
	Rick Lynn	916-356-3612	BAS

Table of Contents

1. Executive Summary

To come. No more than one page – write after rest is done. Follow outline and summarize to conclusions.

2. Introduction

In Q4 2001, Intel employees were interviewed and asked to consider how they use IT. They were asked to categorize IT products and services and explain which ones they used the most often. Their methods for completing IT related tasks were captured and their thoughts about the IT department and the implementation of technology recorded.

As a result of these interviews, the following conclusions were reached about IT on the Web (ITOW):

- There is an inconsistent look & feel throughout the site
- It imposes inefficient navigation for the Intel employee
- It has redundant order entry applications.
- It contains a large amount of outdated content.
- Navigation needs to be usage based as opposed to department based.
- Users want the services that they use the most often on top of the navigation structure.

Subsequently, a recommendation was made for IT to implement content management processes, procedures, and tools for all externally-facing content. BAS was tasked with surveying the Content Management landscape, evaluating the top vendors and providing a recommendation on Content Management for ITOW. The result of that research is this White Paper. It will:

- Provide a primer on Content Management – its origins, components and advantages.
- Discuss how Content Management can positively influence the IT Storefront, and how it will work with the other Web technologies being implemented at Intel.
- Evaluate the three Content Management Software solutions available to Intel at this time: Interwoven TeamSite, Documentum WebPublisher, and Microsoft Content Management Server.
- Provide a recommendation for an Enterprise Content Management solution for IT, which encompasses IT Storefront.

3. Content Management Overview

In this section, we will discuss Content Management – its origins, what it is, what it is NOT, what it involves, how it changes the way an organization works, and how it makes for a more effective workforce.

3.1. Origins of Content Management

Content Management is a technology that has its history embedded within it. An understanding of the technology's origins is required to better understand the technology's shifting design focus and business importance.

There is a story in the aerospace industry that the documentation for a Boeing 747 requires a 747 to transport it from place to place. The origins of content management lie in the need for companies to manage these large amounts of documentation and data required for government regulation, such as in the pharmaceutical and aerospace industries.

Word processors were not enough to manage this level of documentation, because they were designed to enable them to format a fancy document but did extremely little to enable companies to organize and access that information. It was the structure, not the format of the information that was important to such companies

In order to manage this level of documentation, some gave up on word processors and turned to databases. Others invented the *SGML* (Standardized General Markup Language). SGML could create elaborate schemas that, if imposed on a document, could turn it into well-structured data. Unfortunately, SGML was difficult to use, and hard to enforce (except in government-regulated industries).

Document management and source code control systems later took hold because corporate departments and development groups recognized that there was value in providing some control over corporate information and source code, respectively. While these systems provided easier access to such documents, information within documents (especially word processing ones) was still subject only to formatting and not to structural organization.

With the advent of the Web, a paradigm shift began to occur. Content could be something other than documents, specifically Web pages for both public consumption and for the corporate Intranet.

The principal contribution of the Web was to make the concept of a *markup language* acceptable to end users. In fact, HTML - while simplified and only related to display rather than structure - was a direct descendant of SGML. These web pages also needed to be controlled, but managing a web site is quite different from managing internal documents and code:

- *Text, pictures and code must be managed together*, ensuring that the proper elements are all synchronized to deliver a complete web page.
- *Web pages are dynamic rather than static.* Unlike a Word document, many Web pages are created on the fly, combining a static style template with dynamic content and application code.
- *Multiple contributors create a Web site.* The complexity of a Web site means that multiple designers, developers and content contributors must collaborate to create it. In contrast, a single person usually creates an Office document.

Coordination and control of the Web development process fell, by default, to a patchwork of Web masters and developers, creating a digital Balkans of site design and development methods, creating a bottleneck when those people were later tasked to other duties.

3.2. What is Enterprise Content Management?

It is generally accepted that around two thirds of the value of the average corporation is associated with knowledge capital and the value of information. This leaves a third that can be identified as hard assets (money, buildings, machinery, etc.). Intel has had no problem at all investing in applications and systems that manage these hard assets. Content management provides an opportunity to invest in systems and technology that address the far more significant asset we call knowledge and information.

The overall goal of Enterprise Content Management is nothing less than the conversion of content into capital – the value employees and customers place on Intel's business. At a more mundane level, content management addresses the mechanics of storing, archiving, categorizing, securing, distributing, and exploiting content. These mechanisms have to be in place before anything else can happen, and CM provides a broad range of technologies to achieve this.

The common perception - that content management is all about Web content - is misdirected. While the Internet may have been a catalyst to get content on the radar, content includes workgroup documents, rich media, and transactional data, as well as Web content. Content is content, no matter how it is created, and a holistic view offers far greater rewards than a pure focus on Web content. There are as many definitions of Content Management as there are companies selling CM software. But in general terms, Content Management is a process - a way of looking at a business, at the

information that a company uses every day to *do* that business, and finding new ways to use, re-use, recycle and publish that information in a way that:

- Is accurate
- Has a quick time-to-market
- Can be changed quickly with low technical skill
- Can be easily tagged and categorized with metadata
- Can be effectively searched and found by end users
- Can be easily retired when it is no longer relevant
- Provides centralized control of user experience
- Enforces a workflow for the production of content
- Provides multiple output publication from the same information
- Advances the business goals of the company

It is very easy to get distracted by the claims of Content Management System (CMS) vendors about what one product can do vs. the competition. Regardless of which application (or group of applications) is selected, the most important element is the corporate commitment to changing the way we create, manage and distribute our information.

What is also important to note is that the amount of effort (and budget) required to purchase, configure and launch the software and servers is insignificant compared to the effort (and budget) required to implement and maintain such a system. This is not to say, however, that there is not significant ROI to be gained from the installation of a Content Management system.

3.3. What Content Management is not

Enterprise Content Management encompasses many disciplines, and is frequently mistaken for some of its components.

ECM is not Document Management - You can think of document management systems as the file system that you wish your computer had. Many of the functions of DM are also mirrored in CM, but:

- *Document management systems deal primarily with binary files* (in proprietary formats such as MS Word, Excel, PowerPoint, Adobe Acrobat, etc.), while content management systems deal with content components (i.e. ASCII, XML, GIF, JPG, database records, etc.).
- *Document management was invented to manage files that other applications create.* These systems make no attempt to open any of the files under their control and work with what's inside the files.
- *Document management systems exist to provide access to the files under their control.* The purpose of CM is to create publications that are a combination of the components under its control.

ECM is not a Portal - A CMS cannot substitute for efficient portal software such as TIBCO or Plumtree. What IS essential is that it can act as a framework within which a portal application can function. To do this, a CMS can provide essential structural components such as presentation templates and content components such as news articles, content briefs, etc. in a format that can be used by the portal software.

ECM is not a Web Development Tool - ECM is a tool for managing content, not creating it, and should not be expected to provide, or lock developers into, a single tool for content creation. While a CMS may come with tools that allow for the building of web content and integrating it into the repository, it should allow alternative methods –

enabling developers to choose the right tool for the right job, and allowing the adoption of new tools as they become available - without getting in the way.

Perhaps the most common misconception of Content Management is that it is limited to the Web. Theoretically, Content Management enables the production and dissemination of email, printed material and binary files in proprietary formats as well as websites. Content Management is Information Management.

3.4. ECM, Information Management and the SDLC

Information has a lifecycle. When new information is received or developed regarding a product, service or organization, it is necessary to get that information out of the local environment and disseminate it quickly and efficiently. Once disseminated, it is used, reviewed and occasionally altered to meet user needs until it becomes obsolete, at which time it must be retired and archived for reference purposes. This information lifecycle is similar to the Software Development Life Cycle (SDLC).

Content Management brings the SDLC into the world of corporate information management. It provides a **structure** for creating, disseminating, using, changing and retiring corporate information in a controlled system that supplants the traditional ways in which information is made available within the company.

Information within Intel has traditionally been siloed within organizations and business groups. Each group creates information in proprietary formats such as Word and PowerPoint. Sometimes these files are submitted to IT DocLib and categorized, but most is left abandoned once initially viewed. Some is turned over to a developer or a web publisher who creates a web site with the same information.

Each developer has his or her own way of doing things, using different software, methods and skills to put up a site. Each organization has different methods of tracking the information it publishes, different ways of updating that information and different policies for retirement and archiving of obsolete information. Each developer applies his own keywords/metadata, but with little consistency. In short, the Intel Intranet is barely organized chaos.

3.5. ECM and Business Process Reengineering

The traditional publication development cycle starts with business analysts and marketing experts developing content that needs to be published, often in a number of forms – as a report to management in MS Word, as a presentation in MS PowerPoint, as a marketing brochure in Quark Xpress, much of which may or may not end up as part of an internal or external corporate website. Sometimes this content is reused in each of its forms, but frequently a great deal of time is wasted recreating/repurposing this content for each of the media.

Content Management requires that all of these traditional roles be redefined. In a Content Managed web publication system:

- Business analysts, SMEs and marketing specialists create content by putting it into the CMS through data capture templates, which structure the data at a much more granular level than the page level, and allow it to be tagged with metadata.
- This information is passed to editors and metadata experts who confirm that the style is within corporate standards and who apply further metadata, enabling search and reuse. They may also act as approvers, as defined below.
- Approvers review material and determine the form of the output by selecting from multiple output templates
- Such templates can then be scheduled for release, retirement and archiving, all within a single software tool and from a single interface.
- Multiple levels of approvers can be specified for content, based on predetermined permissions and assigned roles.

- All of this workflow routing is accomplished automatically via email.

3.6. Advantages of Content Management

Content management provides a framework for the authoring, design, approval, reuse, categorization and personalization of content. This framework enables the minimum number of support personnel (developers, taxonomists, webmasters and editors) to support the maximum number of content contributors (business analysts, authors and media creators) through:

3.6.1. Centralized Management & Design Control

Content Management systems enable the consolidation of web design and programming into a small group of experts, which:

- Enables tight control of human interface, navigation and look & feel for all Intranet content, providing a consistent user experience
- Enforces programming standards, methods and lifecycle for web pages, web applications and development environments
- Shortens the time to market by optimizing development resources
- Reduces training expenses for new development tools
- Eliminates inexperienced personnel from the web development cycle
- Allows business groups to focus on what they know, not on web development

3.6.2. Distributed Authoring & Approval

CM systems allow large amounts of content to be created, edited, approved and published with no need for any knowledge of HTML or ASP. These systems provide:

- Predetermined user roles: author, editor, approver and publisher. Each of these roles can be assigned to individuals or groups based on NT user names, groups or custom configurations
- Pre-defined, web-based templates for content collection, which restrict contributors to specific, uniform document types.
- Hooks within MS Office tools such as Word which allow templates to be defined which can be submitted directly to the CMS by authors
- Metadata tools that allow editors to apply keywords, taxonomies and search categories to the content
- Configurable approval workflows, allowing each template to be assigned to individuals or groups for approval, editing and publication, and automatically sending emails with access links to them

3.6.3. Reuse & Repurposing of Content

The guiding principal of CM - with apologies to Sun Microsystems - is: “write once, publish anywhere”. CM accomplishes this by enabling a single piece of content (such as a press release) to be assigned to multiple output templates:

- A MS Word template, which will be printed, merged with a mail merge data file, and distributed to the media who require printed releases.
- A fax template, which will be generated and automatically faxed to other media who have requested faxed releases.
- An email template, for automatic emailing to media.

- A corporate press release Web template, which will be automatically scheduled for the external website
- A corporate news template, which will be posted to the company Intranet.
- A syndicated XML file template, which will be pushed to subscribing news sources such as AP, Reuters, Bloomberg, etc.

This reuse of content can save enormous amounts of time and effort over the traditional way, where each use of the content requires separate cutting & pasting, approvals, etc. in a serial rather than parallel workflow.

3.6.4. Resource Conservation and Scalability

Reduces the amount of hardware, support and administration costs required for Intranet presence by enabling multiple groups to share server space more efficiently.

- Consolidating web hosting on fewer servers – for example, when GE moved large parts of their Intranet to a CM solution, they reduced their server count from 14 servers with differing configurations to 4 standardized systems with CMS.
- More efficient use of personnel – fewer servers means less technical support costs
- Ease of scalability – most CM systems are built with scalability in mind, making it much easier to grow a system with fewer problems over the long haul.

3.6.5. Metadata Management & Search

A primary problem with the Intel Intranet is the lack of any organizing structure of content. Many CMS provide tools for the categorization and tagging of content for search and retrieval.

- Automatic workflows that enforce the attachment of standardized metadata to content.
- Metadata management tools that allow for the creation of taxonomies directly related to Intel's information needs, and the ability to assign those taxonomies and keywords to content.
- In a CMS, the metadata gets applied at a much finer granularity than with a document management system, with individual components being categorized and tagged,

3.6.6. Personalization

The metadata that gets applied to the smaller content components enables personalization and the "pushing" of data to the appropriate users.

- Personalization profiles can control the display of content to users based on personal preferences or corporate requirements
- Not all CM systems have personalization built in – what they offer is the framework of metadata with which a dedicated personalization engine can build profiles and deliver content.

3.6.7. Knowledge Management

Content Management acts as a base for true enterprise-level Knowledge Management, where every piece of information within a company is posted, tagged with metadata, indexed, and made available to people who need it. In a true KM system, in addition to traditional content-managed content, even

emails, memos, calendars, and working documents become searchable and available for internal use.

4. Product Evaluations

4.1. Methodology

It was important for us to achieve an objective view of the competing products and not rely strictly on vendor promises and sales pitches to gather product capabilities. We based our evaluations on four m

4.1.1. Research

In order to get a balanced view, we investigated the websites of many CMS vendors. We studied scores of whitepapers, case studies and reports produced by both the vendors and independent analysis firms such as The Gartner Group and Forrester Research. We participated in the *cms-list* online discussion group, gathering impressions from both users and developers of CM systems, both large and small. We referenced *Content Management Bible* by Bob Boiko, which gives a thorough overview of every facet of CM. We also contacted users of CMS within Intel, including NGIP and IOS, and solicited their opinions of existing installations, and the potential for Intranet-wide scalability.

4.1.2. Vendor Interaction & Demonstrations

Once we had narrowed our field down to three major vendors, we contacted each through their respective Intel product managers and requested product demonstrations. We also requested access to technical resources such as product documentation and training materials.

4.1.3. CMS Evaluation Questionnaire and Scoring Matrix

With a lack of clear requirements, we felt it was necessary to cover all of the bases regarding content management.

Using a series of questions in Chapter 16 of Boiko's book, we prepared a 200+ question evaluation matrix covering every aspect of content management. This questionnaire was submitted to vendors - with mixed results, which we felt reflected each individual vendor's desire and commitment to work with us toward an enterprise-wide solution.

Vendor responses were placed in the matrix, and responses were scored on a 0 to 5 scale:

- 0 = Does not meet requirements/no response;
- 1 = Meets requirements with extensive/difficult customization;
- 3 = Meets requirements with relatively simple customization;
- 5 = Meets requirements out-of-the-box

Relative weights to questions based on our perceived importance to IT Storefront and the Intel Intranet on a scale of 0 to 10, with 0 being not relevant to 10 being absolutely required.

4.1.4. Hands-on Product Evaluations

It was especially important to us that we get hands-on experience working with the different tools that come with each product, so that we could get a real-life feel for how the products worked.

To that end, ach product was installed on a base Dell Poweredge 4350/500 Dual Processor Server with 512MB RAM, a 25GB HD and MS Windows 2000 SP3 (Intel Standard Server Build).

Interwoven – Complete

MS – cancelled based on directive from PM

Documentum - underway

4.2. CMS Comparisons

4.2.1. Setup & Configuration

MS – pure play Windows – easiest setup, very resource intensive requires very heavy load on development stations

Interwoven – original UNIX roots show pretty thoroughly – lots of config files, etc. but very light load on development stations

Documentum – under way

4.2.2. Repository Design

Perhaps the most profound difference between the three products is I their approach to repository design.

MS – Pure play SQL Server repository hides XML and other resources within the DBMS infrastructure

Interwoven – pure play XML – all content is stored as XML (except, of course, graphics). Repository is based on a virtual file structure maintained within the CMS that manages checkin/checkout, versioning, etc.

Documentum – stores all as native file format – HTML as HTML, XML as XML, Word as Word, etc., revealing its document management roots

5. Recommendations

1. Don't go it alone. CMS systems are extremely complex and are difficult to implement without help. Regardless of which CMS is chosen, Intel should commit to having extensive help and support from either the vendor or a third-party consulting group that specializes in content management solutions.

Considerably more to come, including final product evaluation.