Managing Unstructured Content in Enterprise Information Portals (EIPs):

Documents, Taxonomies, Content and Knowledge Management in EIPs

May 2001
Table of Contents

Overview of Unstructured Data in Enterprise Portals................................................................. 3

Introduction .................................................................................................................................. 3

The Unstructured Problem ........................................................................................................ 3

Why Organize This Mess in the First Place? ........................................................................... 4

Why is it so Difficult to Organize this Information? ............................................................. 6

The Content Directory in Enterprise Portals ........................................................................... 6

Portal vs. Web Content Management ....................................................................................... 7

Portal Content Management vs. Knowledge Management..................................................... 8

How Does it All Fit Together? .................................................................................................. 9

All About Taxonomies ................................................................................................................. 9

What Exactly is a Taxonomy? .................................................................................................... 9

Automatically Generated Taxonomies? .................................................................................. 10

Industry Standard Taxonomies? ............................................................................................ 11

How Do You Tie Structured and Unstructured Information Together? ................................. 11

Building Taxonomies in EIPs .................................................................................................. 12

The Importance of a Content Management Process ............................................................ 12

Steps for Building a Taxonomy ............................................................................................... 13

Guidelines for Building a Taxonomy – Pitfalls to Avoid ....................................................... 14

Conclusion .................................................................................................................................. 17

About Unitas .............................................................................................................................. 18

Survey ....................................................................................................................................... 18
Overview of Unstructured Data in Enterprise Portals

Introduction
The launch of relational databases and data warehousing led to a revolution in the ability of corporations to collect, organize, query and manage information about their customers, prospects, products and financials, often referred to as structured information.

Today, a new class of software, Enterprise Information Portals ("EIPs") promises to bring order to this corporate jungle of information – often referred to as unstructured information - which knowledge workers need to perform their jobs. A recent estimate put unstructured information at 80% of all corporate information.

This white paper focuses on the issues related to the management, organization, contribution, searching and indexing of unstructured content through the use of a content directory and an organizational Taxonomy as part of an EIP solution.

The Unstructured Problem
The reason that it is hard to peg exactly how much unstructured information a corporation may have is that the information is spread out across hundreds, perhaps even thousands, of individual content stores. First, we offer a broad definition of what we mean by unstructured information:

Unstructured Information – information that contains a significant amount of free form text and does not reside in a structured database.

If this definition seems a bit vague – it’s meant to be. Unstructured information has traditionally been stored as individual documents on local hard drives or on file servers, or as email systems and consist of memos, reports, letters, white papers, marketing material, research, presentations, etc.

Examples of Unstructured Information within a company include:

- **Intranet Sites** – In the past several years, intranet sites have cropped up at an alarming rate in most corporations. While these sites provide significant value to the group that published them, the proliferation of these sites has resulted in inconsistent look and feel, outdated sites, duplication of effort, and most importantly, an inability to navigate these sites in an efficient manner.

- **Email Messages** – Email has become a normal channel of communication. The knowledge that is built up in email threads usually contains important background information on almost every conceivable issue that is being worked on in the company – from sales to back office. More importantly, email from customers and partners is considered valid corporate communications and usually contains key pieces of information about the customer’s or partner’s relationship with the company.
• **Groupware Applications, such as Lotus Notes or Microsoft Exchange Public Folders** – Many companies have used groupware as a means for centralizing information relevant to a particular group or department. While this approach solved one problem, the proliferation of these applications has created another problem – too much information.

• **Public Web Sites** – Companies usually put important information on their public web sites that is important for both internal and external constituents to be kept abreast of – including press releases, news, product information, etc.

• **Extranet Sites** – Secure web sites or extranets are being utilized by more companies as a primary means of communication with their key suppliers, vendors, channel partners and customers. These extranet sites have become another silo of important corporate information. Similarly, many companies subscribe to secure extranets from other companies such as Lexis/Nexis, Bloomberg, Reuters, etc., which provide them with valuable news and research information.

• **Word Processing Documents, Spreadsheets and Presentations** – Desktop files such as Microsoft Office documents are being produced by knowledge workers in all parts of the organization. These documents usually reside on individual hard drives which makes them vulnerable to being lost or on shared file servers - making them difficult to locate. Furthermore, a significant amount of published information is now available in PDF formatted files.

• **Multimedia files and objects** – Many companies have invested in creating graphics, videos, audio files, brochures, scanned images of documents and training materials that exist as multimedia objects on file servers or in media databases. These objects are often spread out among the organization and are not readily available.

Of course, unstructured information isn’t really unstructured. A spreadsheet for example, has structure (columns and rows), as does a memo or an email (a subject field, a from field and a body field). Most text documents have a format or a structure as well; however, for purposes of this paper we will refer to the above definition and examples as unstructured content.

**Why Organize This Mess in the First Place?**

At first glance, the task of managing the all of the unstructured content within an organization (as well as unstructured content from outside the organization) might appear a little daunting. Why organize it in the first place?

We believe that organizations that make the effort to organize their unstructured content into a unified directory, along with a well-defined Content Management process, will reap considerable benefits:
Finding the right information at the right time - The first reason to organize unstructured information is that employees are spending a significant amount of time trying to track down the right piece of information. Often, finding the right piece of information may mean looking at multiple on-line sources, contacting one or more people, possibly accessing multiple business applications and then (hopefully) getting the correct information. This translates into lost employee productivity and higher costs.

It costs too much to reinvent the wheel each time - Faced with not being able to find the right information, employees will often reproduce elements that have already been created in another part of the company. In the long run, this can be of significant cost to organizations - costs that can be saved with an effective Knowledge Portal.

Knowledge walks out the door when someone leaves - Another specific benefit in capturing unstructured content is related to the fact that every time an employee leaves, their associated knowledge and expertise, which companies have invested significant dollars into, leaves with the employee. An effective knowledge capture mechanism can help other employees leverage the knowledge that the former employee has accumulated and can be used by new employees to shorten their ramp-up time.

Reduced Printing Costs and Access Up-to-Date Information - Many companies send out a significant number of paper documents regularly to their employees, customers and business partners, including policies, newsletters, pricing, marketing brochures and manuals, etc. Effectively managing unstructured content on-line can provide employees and customers instant access to the most up-to-date information while saving potentially millions of dollars in annual printing costs.

Improve Partner and Customer Relationships - An effective mechanism for organizing sales, marketing and product information for business partners can lead to several benefits: better relationships with partners, more informed partners and more revenue because of the ability to share best practices. Similarly, customer-facing Portals can increase the responsiveness of the company to customers requests through customer self-service and increase the ability to up-sell customers. By bringing valuable content to the Portal, companies can develop a deeper relationship with their key customers and partners.

Employee Benefits - Effectively managing employee communications and making them available through an Enterprise Portal can help in keeping employees up-to-date on company strategy, company news, changes in policy and upcoming events. It can also help in employee training by capturing what is available in training sessions and making it available through a Portal. These are only some of the business benefits that can be gained by effectively managing unstructured information.
Why is it so Difficult to Organize this Information?

In looking at both the nature of unstructured information and the benefits that could be achieved by organizing it, this leads us to ask the question “Why haven’t more companies made headway in organizing this information?”.

There are a number of key problems that have prevented organizations from being able to fully leverage their unstructured assets:

1. **The information is in different physical locations and stored in different physical and logical formats.** Unlike a database, which collects information into one physical file, unstructured content within a company can reside on tens of thousands of different machines – file servers, email servers, individual desktops, laptop machines and intranet web servers – as well as outside the corporation. There is often a physical access issue in performing a content inventory. Furthermore, the content is usually in a number of different formats, making it difficult to easily consolidate.

2. **The lack of structure makes it difficult to organize content.** By their very nature, unstructured documents are hard to categorize. This is because they often don’t have effective metadata associated with them. For example, a proposal written to IBM doesn’t have a field that says “Client Name: IBM” which would let us categorize it into an IBM folder. In the recent past, some companies have begun to use full text indexing as a way of indexing documents. However, this hasn’t been overly effective as a full text search might return too many documents, making it difficult to get to the information that a user is looking for.

3. **Sometimes information is sitting in people’s heads and not available in electronic format.** Another problem with organizing unstructured information is that it is often not written down, but resides in employee’s heads as pure “knowledge”. Enterprise Portals can make it easier to access this information with “expert locators” which direct you to the right person to answer a specific question.

4. **The time and manpower needed to categorize.** Many Internet Portals (like Yahoo!) have a team of people that categorize web sites and ensure that they fit into the appropriate directory or subdirectory. Many corporations may not have enough manpower to spare to perform the manual categorization.

The Content Directory in Enterprise Portals

The term “Content Management” has been used in several ways in the industry. This in turn has caused confusion about the type of Content Management available within most EIP products.

Most Enterprise Information Portal solutions have a method to organize the collection, management and distribution of unstructured information – it is generically referred to as a Taxonomy. Taxonomies have many subtitles to them – some of which we will describe in this white paper, but it’s easiest to think of a Taxonomy as an Enterprise Table of Contents or a Content Directory.
The best way to understand a Content Directory is to think of it as a collection of channels or folders. Each of these channels is a logical view of the underlying data and each channel can contain unstructured content (documents, web pages, etc.) or sub-channels.

If you have ever used Yahoo! or another Internet search engine, you’re intuitively familiar with a Taxonomy. Figure 1 shows an overview of the Taxonomy used by Yahoo for organizing the content of the World Wide Web. Yahoo’s channels are broad categories such as business or entertainment, which are then narrowed down into sub-channels such as finance, entrepreneurship, etc.

Most EIPs have functionality for the creation and maintenance of a Taxonomy; however, they go to varying degrees of depth in doing so. There are a number of features, which are related to the management of unstructured content within an EIP environment. The Unitas Portal evaluation methodology rates each of the EIP Portal vendors in these areas.

**Portal vs. Web Content Management**

A source of confusion in the Enterprise Portal space has been the term “Content Management”. In order to clear up the confusion we have to draw a distinction between “Web Content Management” and “Portal Content Management”.

“Web Content Management” solutions, such as those available from companies like Interwoven and Vignette, provide a way for a diverse group of people to edit web pages...
and submit them for approval. In this model, the product contains “templates” of the type of pages that are on the web site (“Press Releases”). The layouts of all of the press releases are the same – only the text is different. End users who have the proper privileges can log in and, via a web browser, edit the text on the web page, but they cannot change the layout of the page.

“Content Management” within a document management system usually involves the submission of individual documents into folders or directories, which can then go into an approval cycle. This is much closer to the process of Content Management within Enterprise Portals. A Content Directory contains multiple folders. Each of the folders contains documents (of any type) or links to web pages. Within the context of an Enterprise Portal, Content Management refers to the indexing, maintenance and expiration of documents and links that reside within the folders of the Content Directory.

So, Content Management for the Portal does not usually involve the editing of pages on the web site, but rather the organization of those pages into a Content Directory. If you wanted to edit the pages, you would use a web page editor or a Web Content Management system. Similarly, if a user wanted to edit a Microsoft Word document within the Content Directory, the user would need to use an editor like Microsoft Word.

**Portal Content Management vs. Knowledge Management**

Another popular term that gets confused with Enterprise Portals is Knowledge Management (or KM). KM is often confused with EIPs because the goal of an EIP is often the same as the goal of Knowledge Management: to harvest the knowledge that has been produced by the organization and make it available to other employees.

The argument has been made that the most important asset a corporation has is the intellectual capital that its employees have accumulated over the corporation’s existence. While some of this knowledge may be encapsulated in hard assets such as methodologies, software or manuals, much of it often exists as soft assets - information gathered during meetings, knowledge of processes that aren’t documented or emails that have been exchanged between employees. Knowledge Management is also a term that lacks a clear definition, but it generally refers to the collection, categorization, searching, distribution and on-going management of information assets.

Most EIP products have features that help with Knowledge Management – an expert locator, for example, helps to find the resident experts within the corporation. Instant messaging and collaboration features can help to connect with the expert (who might be in another part of the country) and leverage their knowledge base. Similarly, employees can generally contribute knowledge to the Portal by adding documents, participating in forums and generally making themselves available as knowledge experts.

Portal Content Management can be seen as part of an overall Knowledge Management strategy. Within the Portal, Content Management usually refers to the organizing of unstructured assets into a content directory and the ongoing maintenance of those documents.
How Does it All Fit Together?

Figure 2: How does all the unstructured content fit together?

Figure 2 shows how all of the different aspects of unstructured data fit together into a comprehensive EIP solution. Across the right hand side are the most common features that EIPs provide for finding and managing unstructured data. This includes searching a common index across many different physical content sources and browsing the directory which organizes the content into folders and subfolders. Web Content Management usually provides web site authoring and sits behind the intranet or web sites, while Portals have the ability to deposit and approve content (documents and links to web pages) within the EIP. Knowledge Management is usually implemented within the Portal via the Content Management process and features such as contribution of documents, adding suggestions and expert locators.

All About Taxonomies

What Exactly is a Taxonomy?

Although a Taxonomy can be defined simply as the content directory for an enterprise’s unstructured information, it can be populated with content and presented to the user in many different ways. Both substance and behavior of the directory define usable Taxonomy for an enterprise.

Just as the table of contents for a book might include entries for the book’s parts, sections, chapters and topics, the Taxonomy for a business Portal organizes content of interest to the business in a similar structure. People who use PCs are familiar with the way content is stored on their computer’s hard disk and network file servers. This
content is usually displayed to the user in a graphical interface that uses images that look like small file folders. Some document management systems go further with the physical file analogy and present folders in images that represent file cabinets and file rooms. These metaphors help people leverage their familiarity with the organization of paper files to adapt to the electronic storage and access methods. In reality, of course, the physical storage of electronic content has nothing to do with icons representing the fixed asset facilities that preceded computer technology.

Indented lists, classification trees and hierarchies are other terms used to describe Taxonomy structures. Folders and sub-folders, topics and sub-topics, categories and sub-categories are others. Regardless of the terms used to describe a Taxonomy, however, it gives us a way to organize content into a structure that is easily browsed by the Portal user. Each node or element in the structure corresponds to a logical collection of content that is stored anywhere within the enterprise’s electronic media devices or accessible over the Internet.

Products like Autonomy have other graphical presentations available for users. Consider also the Internet site called WebBrain (http://www.webbrain.com), which presents a Taxonomy of World Wide Web content as displayed in the following image:

![Figure 3: Alternative Presentation of World Wide Web Taxonomy](image)

Compare this presentation to the image of the Yahoo web site (Figure 1) in the previous section of the paper. Both web sites provide a way to browse the same content on the World Wide Web, but the browsing metaphors used are completely different.

**Automatically Generated Taxonomies?**

With the importance of easy and timely access to a corporation’s unstructured data increasing as rapidly as the volume of content itself, many approaching this problem will consider techniques and products to generate their Taxonomies automatically. The good news is that technologies are improving and the results are beginning to show business benefit. Products from Autonomy, Semio, Lotus and 80/20 Software are notable in this
market - although each product has its own approach and proprietary methods for categorizing content.

Taxonomy automation can be particularly useful to help organize large volumes of corporate data and to help keep new content organized. The key word here is “help,” since some manual effort to improve the jump-start achieved through automation is almost always necessary. End-user input into the process from inception of the system and over time as the application is used can greatly improve the efficacy of automation and the value provided to the organization.

In the business world, Taxonomies are more of an art than a science. The more diverse the business areas of an organization and its users, the greater the complexity of organizing content into categories that meet the needs in a single global Taxonomy. Debates can go on endlessly and a tool can enlighten the process and facilitate initial results. As with any automated “package” solution, however, there are trade-offs to consider between letting the tool drive the business and training the tool to deliver targeted strategic elements of the overall solution.

Industry Standard Taxonomies?
Business, government and industry are categorized in many Taxonomies maintained by standards associations. The North American Industry Classification System (NAICS), the Classifications of the Functions of Government (COFOG) and Standard Industry Classification (SIC) are examples. More specific Taxonomies for particular business types and industry segments are also available. Such reference Taxonomies can be useful for classifying some of an enterprise’s content, but they will not provide the company-specific categories that are needed to support the knowledge assets of an entire enterprise.

Standard Taxonomies in use by business and industry will, however, help guide an enterprise document classification effort if only for the ideas and discussions they trigger. Regardless of whether a manual or automated approach is used the effort should include analysis and design sessions a rigorous survey of existing content and knowledge of how workers interact with the content to carry out the business of the enterprise.

How Do You Tie Structured and Unstructured Information Together?
By definition, Enterprise Information Portals provide a single point of access to internal business systems and information content, both inside and outside a company’s boundaries. Out-of-the-box modules included with commercial EIP package software facilitate connecting the Portal to the various back-end systems and databases. What is much more challenging, however, is the integration of information from a diverse set of sources to support a business process. While easy desktop access to content from separate sources is clearly attractive to a work force familiar with scavenging for information assets, much more value comes from the targeted delivery of relevant information from both structured and unstructured data sources.
How this is accomplished in the Portal is the subject of another Unitas white paper, “Bringing Structured and Unstructured Information Together in an EIP”.

Building Taxonomies in EIPs

The Importance of a Content Management Process

The single most important aspect of an EIP implementation is establishing a Content Management process for the Portal. Providing easy access to corporate information from a variety of sources – both structured and unstructured – is the primary benefit of the Portal to its users. Without a process to collect, organize and present its content efficiently and diligently, the value of the Portal is greatly diminished.

The Pitfalls section of this paper touches on many pros and cons to Portal implementation approaches, including several related to Content Management. The main things to consider in establishing a Content Management process are:

- **Content segmentation and access control** – Who can deposit, read and delete documents? Who can create additional folders and otherwise modify Taxonomy? Who creates and manages the content crawlers and indexing jobs to make remote content available for browsing and searching? Who can add notification alerts and subscribe to content folders? Which users should have the capability to save personal searches in the Portal?

- **Archival process** – What do you do with old content? Who determines when content has out-lived its usefulness and should be retired from the Portal?

- **Contribution and approval process** – How do knowledge workers of the enterprise contribute content to be shared with others? What incentives are provided to encourage submission of assets to content areas of the Taxonomy lacking substance and variety?

- **Distributed vs. centralized** – Content can be managed successfully by either a core central team serving the entire enterprise or by a distributed content workforce. When do you choose one approach over the other?

- **Manual vs. automated** – When are products that automate Taxonomy creation appropriate for an enterprise? What’s the best manual approach for content categorization?

In defining a Content Management approach for an enterprise it might be worth examining the Open Directory Project on the Internet. “As the web grows, automated search engines and directories with small editorial staffs will be unable to cope with the volume of sites,” its overview says. “The Open Directory Project's goal is to produce the most comprehensive directory of the web, by relying on a vast army of volunteer editors.” Their Home Page is shown in Figure 4: Open Directory Project. Compare this to Yahoo’s directory (Figure 1) of the same Internet content. Many of the same categories appear at the top of the directory structures. Yahoo’s directory is managed by
a large paid staff. The Open Directory Project has over 366,000 categories organized by a volunteer Content Management community of more than 36,000 editors.

Steps for Building a Taxonomy

Building a Taxonomy can be a simple or a very involved process. In the worst case, it has been compared to “black magic”. In the best case, it has been compared to the ordered process of implementing a data model for an enterprise data warehouse. The reality is somewhere in between. It is possible to enumerate the steps of building a Taxonomy; however, each step requires a certain amount of subjective input, usually gained from experience in building Taxonomies.

As part of our D4 Portal Deployment Methodology, Unitas has a detailed set of steps for defining, building and refining a Taxonomy. Although we do not present the methodology and its supporting documentation in this white paper, here are a few steps:

- Conduct a content inventory of existing content sources
• Prioritize content sources and conduct detailed analysis of key content sources, including existing metadata and usage analysis

• Create a high level Taxonomy

• Define a Content Management process

• Fill out the details of the Taxonomy folders of the Taxonomy

• Test and refine the content acquisition crawlers

• Test and refine the search strategy

• Refine the Taxonomy based on the results of both the crawling of content and the results of searching the content

• Manage content within folders on an on-going basis in a distributed way

**Guidelines for Building a Taxonomy – Pitfalls to Avoid**

**Not Having a Strategy and Plan**
Organizing the unstructured data assets of an enterprise should be approached like any other initiative requiring an investment of people, time and effort – with a plan. Although portions of the infrastructure and framework for addressing a company’s unstructured data needs may be provided by the enterprise Portal product, the road map for using these components most effectively for a particular company is not.

Planning should start with an examination of the company’s business strategy. Next, determine the extent to which information from unstructured sources contributes to that strategy and plan how the vast amounts of available data can be organized and delivered efficiently to everyone when they need it to perform their job. Strategic alignment of the enterprise information Portal solution, as well as each component of the solution, results in a more effective use of information resources and a greater return on investment.

**Insufficient Cost Benefit Analysis**
Along with a strategy and plan for harnessing actionable information lost in a corporation’s unstructured data, the importance of a cost benefit analysis should not be overlooked. A cost benefit analysis for implementing a solution for managing unstructured content will help build the business case for taking action and facilitate the process of funding the project. Almost as important, however, is the value of this analysis to identifying the highest priority areas of focus and the scope of the effort. In addition, the cost benefit analysis will become a vehicle for communicating the solution to the enterprise and setting expectations for the projected outcome of the investment.

*(Note: Portal cost benefit analysis is the subject of another Unitas White Paper.)*
Trying to “Solve World Hunger”
Start small, but anticipate growth and change. With 80% or more of a company’s data in files of different types, it is not uncommon for large corporations to have terabytes of unstructured data stored on its file servers, PCs and laptops. Efforts that try to organize all of this content at once for the whole enterprise will almost always fail. Unstructured Content Management solutions are major undertakings with volumes this high and successful initiatives will use an iterative approach to delivering results. Managing scope and expectations through a series of iterative releases, starting with a proof of concept has many benefits. Building and deploying the solution iteratively will accomplish some results early, expand the scope and broaden the impact of the solution at a palatable pace to the organization.

Assuming One Solution Fits All
With unstructured information assets there are many Taxonomies that do or could define the logical categories for an enterprise. Such factors as industries, lines of service, geography and culture all influence the way content is organized. It would be a mistake to assume that a single Taxonomy or Content Management approach is appropriate for every Portal implementation. Even within the same company many Taxonomies or variants may be necessary to be completely sensible for the different business areas of the enterprise. Indeed, some large global corporations support the concept of a distinct Taxonomy for each of its business units, communities of interest and project teams. This flexibility requires a large investment in the system architecture design of the Portal as well as a decentralized approach to Content Management.

Inappropriate Use of “Divide and Conquer” Approach
In designing and implementing the content directory for your enterprise, consider an approach where the responsibility for different components of the Taxonomy is distributed across organizations that will use the directory. This is particularly important for large global corporations with separate businesses. Each major group or division is apt to have very distinct content directories. Some large institutions have established a high level Taxonomy corresponding to their lines of business and major organizations. Each organization is given the Content Management responsibility, the tools and support necessary to enable such a distributed approach to be successful. Whenever Content Management is distributed among several content managers, consider sponsoring a community within your Portal where these individuals can collaborate to share their knowledge and experiences.

Relying Too Much on Automated Solutions
Although there are technically advanced products that can build Taxonomies with little or no human involvement, completely automated solutions are not recommended for most enterprises. Such implementations rarely provide the optimal categorization of content for efficient use by the business. A team expecting immediate, usable results from one of these products “out of the box” will soon discover that some manual effort will be required to tailor what the product provides into an appropriate and manageable Taxonomy. Human involvement is usually required.
Ignoring Change Management – “If We build it, They Will come.”
In spite of best efforts to implement a Portal with a well-crafted content directory, it is a mistake to assume all knowledge workers within the enterprise will adapt to the environment immediately. Deployment support for solutions that address unstructured data is as important as it is with any new system. Sharing content that has always been lost with the deletion of an e-mail or stored in personal folders on a laptop, will not start automatically once an enterprise content directory is in place.

Assuming Things are “Filed” Correctly – What Metadata?
Any content directory that relies on document properties for automated classification or the completion of metadata forms with content submissions is bound to fail without the constant attention and scrutiny of the content manager. Contributing metadata is a clear impediment to the submission of the content artifacts themselves. Implementers of Knowledge Management solutions need to consider alternatives to placing the burden of metadata management on a company’s knowledge workers themselves. That said, however, distributed responsibility carried out by a team of content managers dedicated to their mission can provide the level of involvement necessary to ensure content is labeled and categorized for effective classification, search, retrieval and business reuse.

“I Don’t Know what I Don’t Know”
How can you be sure that you are finding all the good content available when you need it to make a business decision? One of the major complaints about Internet searches is that the results frequently include thousands of items, many of them irrelevant to your specific need. The sequence in which the results are presented to you is often influenced by someone else’s objectives instead of your own. If there is useful content in the results list returned from these searches, it may be too far down the list of several thousand items for you to find it. Finally, the most useful content available may not be returned in a search results list at all. Even the best managed environments with good search indices can fall short of delivering the cream of the crop when it’s needed most by the business Portal user. Communicating the directory’s evolving breadth and depth can help manage these expectations. Supporting seekers with viable alternatives when content is not available through the Portal is also recommended.

Lack of an Effective Archival Process
Business content has a lifespan of usefulness. Failing to refresh content and remove content that is no longer appropriate for reuse can create problems as serious as those we work to solve when we implement a Portal. Many factors determine when content should be retired from active use, but the best approach to keeping content vital involves a defined practice of content “brokering” plus automated or manual archival (and renewal) based on asset type and business parameters.

Not Taking Action
With so many possible derailments and ditches threatening successful unstructured data initiatives, if might be tempting to do nothing at all about this problem. With the information value deficit so enormous in most corporations today, however, failure to take action is probably the biggest pitfall to avoid. The benefits of a Portal
implementation and an integral approach for unstructured Content Management far outweigh the investments made to do it right. The Unitas EIP solution includes components to enable our clients to meet every one of these challenges and to celebrate success with each phase of the implementation process and every release of the solution deployment.

**Conclusion**

In most organizations, there is a voluminous amount of unstructured content that often represents the key intellectual assets of the organization. Because of difficulties inherent in unstructured content, many organizations have just now begun to address the management of that content and reap substantial benefits. Enterprise Information Portals provide an ideal platform on which to help employees find and manage unstructured content, through the use of a Content Directory and the ability to automate aspects of a Knowledge Management or Content Management process. Many corporations can achieve important business benefits by going down this path.
About Unitas

Unitas Corporation specializes in the rapidly emerging market of enterprise Portals. As the only IT services company focused specifically on planning, developing, integrating and implementing enterprise Portals, Unitas fuses e-business strategies with enterprise Portal technology solutions to create united business communities within the corporation and among customers, partners and suppliers. The company’s consulting expertise and implementation services help large and mid-size corporations apply their knowledge assets through an enterprise Portal to create strategic advantage and operational value in the networked economy.

Unitas Corporation assists with all aspects of Enterprise Information Portal planning, analysis, integration and deployment. We have out of the box “Evaluation Workshops” and “Discovery Workshops” to help you choose the best solution, get buy-in from senior management and plan an Enterprise Information Portal Solution. We then provide custom services to rapidly develop, integrate and deploy a Portal solution – sometimes in as little as 60 days.

For more information about other Portal white papers or “Portal Buzz”, please contact:

Unitas Corporation
404 Wyman Street, Suite 320
Waltham, MA  02451
Phone:  781.522.9300
Fax:  781.522.9350
www.unitas.com

Survey

Did you find the information in this white paper useful?  Please take a few moments to give us feedback on this white paper and future white papers that would be of interest to you by completing our on-line survey under the ‘white papers’ section of the Unitas web site at www.unitas.com